

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
20 December 2001 (20.12.2001)

PCT

(10) International Publication Number
WO 01/96390 A2

(51) International Patent Classification⁷: C07K 14/47

(21) International Application Number: PCT/US01/18577

(22) International Filing Date: 8 June 2001 (08.06.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/210,821 9 June 2000 (09.06.2000) US
60/256,571 18 December 2000 (18.12.2000) US
60/290,240 10 May 2001 (10.05.2001) US

(71) Applicant (for all designated States except US): CORIXA CORPORATION [US/US]; 1124 Columbia Street, Suite 200, Seattle, WA 98104 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): JIANG, Yuqiu [CN/US]; 5001 S. 232nd Street, Kent, WA 98032 (US). HEPLER, William, T. [US/US]; 12034 38th Avenue N.E., Seattle, WA 98125 (US). CLAPPER, Jonathan, D. [US/US]; 2149 Dexter Avenue N., #4, Seattle, WA 98109 (US). WANG, Aijun [CN/US]; 3106 213th Place S.E., Issaquah, WA 98029 (US). SECRIST, Heather [US/US]; 3844 35th Avenue W., Seattle, WA 98199 (US).

(74) Agents: POTTER, Jane, E., R.; Seed Intellectual Property Law Group PLLC, Suite 6300, 701 Fifth Avenue, Seattle, WA 98104-7092 et al. (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: COMPOSITIONS AND METHODS FOR THE THERAPY AND DIAGNOSIS OF COLON CANCER

(57) Abstract: Compositions and methods for the therapy and diagnosis of cancer, such as colon cancer, are disclosed. Compositions may comprise one or more colon tumor proteins, immunogenic portions thereof, or polynucleotides that encode such portions. Alternatively, a therapeutic composition may comprise an antigen presenting cell that expresses a colon tumor protein, or a T cell that is specific for cells expressing such a protein. Such compositions may be used, for example, for the prevention and treatment of diseases such as colon cancer. Diagnostic methods based on detecting a colon tumor protein, or mRNA encoding such a protein, in a sample are also provided.

WO 01/96390 A2

COMPOSITIONS AND METHODS FOR THE THERAPY AND DIAGNOSIS OF COLON CANCER

TECHNICAL FIELD OF THE INVENTION

The present invention relates generally to therapy and diagnosis of cancer, such as colon cancer. The invention is more specifically related to polypeptides comprising at least a portion of a colon tumor protein, and to polynucleotides encoding such polypeptides. Such polypeptides and polynucleotides may be used in vaccines and pharmaceutical compositions for prevention and treatment of colon malignancies, and for the diagnosis and monitoring of such cancers.

10 BACKGROUND OF THE INVENTION

Cancer is a significant health problem throughout the world. Although advances have been made in detection and therapy of cancer, no vaccine or other universally successful method for prevention or treatment is currently available. Current therapies, which are generally based on a combination of chemotherapy or surgery and radiation, continue to prove inadequate in many patients.

Colon cancer is the second most frequently diagnosed malignancy in the United States as well as the second most common cause of cancer death. The five-year survival rate for patients with colorectal cancer detected in an early localized stage is 92%; unfortunately, only 37% of colorectal cancer is diagnosed at this stage. The survival rate drops to 64% if the cancer is allowed to spread to adjacent organs or lymph nodes, and to 7% in patients with distant metastases.

The prognosis of colon cancer is directly related to the degree of penetration of the tumor through the bowel wall and the presence or absence of nodal involvement, consequently early detection and treatment are especially important. Currently, diagnosis is aided by the use of screening assays for fecal occult blood, sigmoidoscopy, colonoscopy and double contrast barium enemas. Treatment regimens are determined by the type and stage of the cancer, and include surgery, radiation therapy and/or chemotherapy. Recurrence following surgery (the most common form of therapy) is a major problem and is often the ultimate cause of death.

In spite of considerable research into therapies for these and other cancers, colon cancer remains difficult to diagnose and treat effectively. Accordingly, there is a need in the art for improved methods for detecting and treating such cancers. The present invention fulfills these needs and further provides other related advantages.

5 SUMMARY OF THE INVENTION

In one aspect, the present invention provides polynucleotide compositions comprising a sequence selected from the group consisting of:

- (a) sequences provided in SEQ ID NOs:1-234, 236, and 244;
- (b) complements of the sequences provided in SEQ ID NOs:1-234,
10 236, and 244;
- (c) sequences consisting of at least 20, 25, 30, 35, 40, 45, 50, 75 and 100 contiguous residues of a sequence provided in SEQ ID NOs:1-234, 236, and 244;
- (d) sequences that hybridize to a sequence provided in SEQ ID NOs:1-234, 236, and 244, under moderate or highly stringent conditions;
- 15 (e) sequences having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% identity to a sequence of SEQ ID NOs:1-234, 236, and 244;
- (f) degenerate variants of a sequence provided in SEQ ID NOs:1-234, 236, and 244.

20 In one preferred embodiment, the polynucleotide compositions of the invention are expressed in at least about 20%, more preferably in at least about 30%, and most preferably in at least about 50% of colon tumor samples tested, at a level that is at least about 2-fold, preferably at least about 5-fold, and most preferably at least about 10-fold higher than that for normal tissues.

25 The present invention, in another aspect, provides polypeptide compositions comprising an amino acid sequence that is encoded by a polynucleotide sequence described above.

The present invention further provides polypeptide compositions comprising an amino acid sequence selected from the group consisting of sequences
30 recited in SEQ ID NOs:235, 237, and 245.

In certain preferred embodiments, the polypeptides and/or polynucleotides of the present invention are immunogenic, *i.e.*, they are capable of eliciting an immune response, particularly a humoral and/or cellular immune response, as further described herein.

5 The present invention further provides fragments, variants and/or derivatives of the disclosed polypeptide and/or polynucleotide sequences, wherein the fragments, variants and/or derivatives preferably have a level of immunogenic activity of at least about 50%, preferably at least about 70% and more preferably at least about 90% of the level of immunogenic activity of a polypeptide sequence set forth in SEQ
10 ID NOs:235, 237, and 245 or a polypeptide sequence encoded by a polynucleotide sequence set forth in SEQ ID NOs:1-234, 236, and 244.

The present invention further provides polynucleotides that encode a polypeptide described above, expression vectors comprising such polynucleotides and host cells transformed or transfected with such expression vectors.

15 Within other aspects, the present invention provides pharmaceutical compositions comprising a polypeptide or polynucleotide as described above and a physiologically acceptable carrier.

Within a related aspect of the present invention, the pharmaceutical compositions, *e.g.*, vaccine compositions, are provided for prophylactic or therapeutic
20 applications. Such compositions generally comprise an immunogenic polypeptide or polynucleotide of the invention and an immunostimulant, such as an adjuvant.

The present invention further provides pharmaceutical compositions that comprise: (a) an antibody or antigen-binding fragment thereof that specifically binds to a polypeptide of the present invention, or a fragment thereof; and (b) a physiologically
25 acceptable carrier.

Within further aspects, the present invention provides pharmaceutical compositions comprising: (a) an antigen presenting cell that expresses a polypeptide as described above and (b) a pharmaceutically acceptable carrier or excipient. Illustrative antigen presenting cells include dendritic cells, macrophages, monocytes, fibroblasts
30 and B cells.

Within related aspects, pharmaceutical compositions are provided that comprise: (a) an antigen presenting cell that expresses a polypeptide as described above and (b) an immunostimulant.

The present invention further provides, in other aspects, fusion proteins
5 that comprise at least one polypeptide as described above, as well as polynucleotides encoding such fusion proteins, typically in the form of pharmaceutical compositions, e.g., vaccine compositions, comprising a physiologically acceptable carrier and/or an immunostimulant. The fusions proteins may comprise multiple immunogenic polypeptides or portions/variants thereof, as described herein, and may further comprise
10 one or more polypeptide segments for facilitating the expression, purification and/or immunogenicity of the polypeptide(s).

Within further aspects, the present invention provides methods for stimulating an immune response in a patient, preferably a T cell response in a human patient, comprising administering a pharmaceutical composition described herein. The
15 patient may be afflicted with colon cancer, in which case the methods provide treatment for the disease, or patient considered at risk for such a disease may be treated prophylactically.

Within further aspects, the present invention provides methods for inhibiting the development of a cancer in a patient, comprising administering to a
20 patient a pharmaceutical composition as recited above. The patient may be afflicted with colon cancer, in which case the methods provide treatment for the disease, or patient considered at risk for such a disease may be treated prophylactically.

The present invention further provides, within other aspects, methods for removing tumor cells from a biological sample, comprising contacting a biological
25 sample with T cells that specifically react with a polypeptide of the present invention, wherein the step of contacting is performed under conditions and for a time sufficient to permit the removal of cells expressing the protein from the sample.

Within related aspects, methods are provided for inhibiting the development of a cancer in a patient, comprising administering to a patient a biological
30 sample treated as described above.

Methods are further provided, within other aspects, for stimulating and/or expanding T cells specific for a polypeptide of the present invention, comprising contacting T cells with one or more of: (i) a polypeptide as described above; (ii) a polynucleotide encoding such a polypeptide; and/or (iii) an antigen presenting cell that expresses such a polypeptide; under conditions and for a time sufficient to permit the stimulation and/or expansion of T cells. Isolated T cell populations comprising T cells prepared as described above are also provided.

Within further aspects, the present invention provides methods for inhibiting the development of a cancer in a patient, comprising administering to a patient an effective amount of a T cell population as described above.

The present invention further provides methods for inhibiting the development of a cancer in a patient, comprising the steps of: (a) incubating CD4⁺ and/or CD8⁺ T cells isolated from a patient with one or more of: (i) a polypeptide comprising at least an immunogenic portion of polypeptide disclosed herein; (ii) a polynucleotide encoding such a polypeptide; and (iii) an antigen-presenting cell that expressed such a polypeptide; and (b) administering to the patient an effective amount of the proliferated T cells, and thereby inhibiting the development of a cancer in the patient. Proliferated cells may, but need not, be cloned prior to administration to the patient.

Within further aspects, the present invention provides methods for determining the presence or absence of a cancer, preferably a colon cancer, in a patient comprising: (a) contacting a biological sample obtained from a patient with a binding agent that binds to a polypeptide as recited above; (b) detecting in the sample an amount of polypeptide that binds to the binding agent; and (c) comparing the amount of polypeptide with a predetermined cut-off value, and therefrom determining the presence or absence of a cancer in the patient. Within preferred embodiments, the binding agent is an antibody, more preferably a monoclonal antibody.

The present invention also provides, within other aspects, methods for monitoring the progression of a cancer in a patient. Such methods comprise the steps of: (a) contacting a biological sample obtained from a patient at a first point in time with a binding agent that binds to a polypeptide as recited above; (b) detecting in the

sample an amount of polypeptide that binds to the binding agent; (c) repeating steps (a) and (b) using a biological sample obtained from the patient at a subsequent point in time; and (d) comparing the amount of polypeptide detected in step (c) with the amount detected in step (b) and therefrom monitoring the progression of the cancer in the patient.

The present invention further provides, within other aspects, methods for determining the presence or absence of a cancer in a patient, comprising the steps of: (a) contacting a biological sample, e.g., tumor sample, serum sample, etc., obtained from a patient with an oligonucleotide that hybridizes to a polynucleotide that encodes a polypeptide of the present invention; (b) detecting in the sample a level of a polynucleotide, preferably mRNA, that hybridizes to the oligonucleotide; and (c) comparing the level of polynucleotide that hybridizes to the oligonucleotide with a predetermined cut-off value, and therefrom determining the presence or absence of a cancer in the patient. Within certain embodiments, the amount of mRNA is detected via polymerase chain reaction using, for example, at least one oligonucleotide primer that hybridizes to a polynucleotide encoding a polypeptide as recited above, or a complement of such a polynucleotide. Within other embodiments, the amount of mRNA is detected using a hybridization technique, employing an oligonucleotide probe that hybridizes to a polynucleotide that encodes a polypeptide as recited above, or a complement of such a polynucleotide.

In related aspects, methods are provided for monitoring the progression of a cancer in a patient, comprising the steps of: (a) contacting a biological sample obtained from a patient with an oligonucleotide that hybridizes to a polynucleotide that encodes a polypeptide of the present invention; (b) detecting in the sample an amount of a polynucleotide that hybridizes to the oligonucleotide; (c) repeating steps (a) and (b) using a biological sample obtained from the patient at a subsequent point in time; and (d) comparing the amount of polynucleotide detected in step (c) with the amount detected in step (b) and therefrom monitoring the progression of the cancer in the patient.

Within further aspects, the present invention provides antibodies, such as monoclonal antibodies, that bind to a polypeptide as described above, as well as

diagnostic kits comprising such antibodies. Diagnostic kits comprising one or more oligonucleotide probes or primers as described above are also provided.

These and other aspects of the present invention will become apparent upon reference to the following detailed description. All references disclosed herein are
5 hereby incorporated by reference in their entirety as if each was incorporated individually.

BRIEF DESCRIPTION OF THE SEQUENCE IDENTIFIERS

SEQ ID NO: 1 is the determined cDNA sequence for 54172.1.

SEQ ID NO: 2 is the determined cDNA sequence for 54104.1 which
10 shares homology with PAC 75N13 on chromosome Xq21.1.

SEQ ID NO: 3 is the determined cDNA sequence for 53978.1 which shares homology with Glutamine:fructose-6 phosphate amidotransferase.

SEQ ID NO: 4 is the determined cDNA sequence for 54184.1 which shares homology with Colon Kruppel-like factor.

15 SEQ ID NO: 5 is the determined cDNA sequence for 54149.1 which shares homology with cDNA FLJ10461 fis, clone NT2RP1001482.

SEQ ID NO: 6 is the determined cDNA sequence for 54034.1.

SEQ ID NO: 7 is the determined cDNA sequence for 54085.1 which shares homology with Human beta 2 gene.

20 SEQ ID NO: 8 is the determined cDNA sequence for 53948.1 which shares homology with 12p12 BAC RPC111-267J23.

SEQ ID NO: 9 is the determined cDNA sequence for 54026.1 which shares homology with Clone 164F3 on chromosome X2q21.33-23.

25 SEQ ID NO: 10 is the determined cDNA sequence for 53907.1 which shares homology with Lysyl hydroxylase isoform 2.

SEQ ID NO: 11 is the determined cDNA sequence for 54066.1 which shares homology with Mucin 11.

SEQ ID NO: 12 is the determined cDNA sequence for 54017.1 which shares homology with Mucin 11.

SEQ ID NO: 13 is the determined cDNA sequence for 54006.1 which shares homology with Mucin 11.

SEQ ID NO: 14 is the determined cDNA sequence for 53962.1 which shares homology with Epiregulin (EGF family).

5 SEQ ID NO: 15 is the determined cDNA sequence for 54028.1 which shares homology with Mucin 12.

SEQ ID NO: 16 is the determined cDNA sequence for 54166.1 which shares homology with E1A enhancer binding protein.

10 SEQ ID NO: 17 is the determined cDNA sequence for 54174.1 which shares homology with PAC clone RP1-170O19 from 7p15-p21.

SEQ ID NO: 18 is the determined cDNA sequence for 53949.1.

SEQ ID NO: 19 is the determined cDNA sequence for 53898.1.

SEQ ID NO: 20 is the determined cDNA sequence for 54069.1.

15 SEQ ID NO: 21 is the determined cDNA sequence for 54048.1 which shares homology with cDNA FLJ20676 fis, clone KAlA4294.

SEQ ID NO: 22 is the determined cDNA sequence for 54031.1 which shares homology with Chromosome 17, clone hRPC.1171_1_10.

SEQ ID NO: 23 is the determined cDNA sequence for 54154.1 which shares homology with Alpha topoisomerase truncated form.

20 SEQ ID NO: 24 is the determined cDNA sequence for 54009.1 which shares homology with Cytokeratin 20.

SEQ ID NO: 25 is the determined cDNA sequence for 54070.1 which shares homology with Erythroblastosis virus oncogene homolog 2.

25 SEQ ID NO: 26 is the determined cDNA sequence for 53998.1 which shares homology with Polyadenylate binding protein II.

SEQ ID NO: 27 is the determined cDNA sequence for 54089.1.

SEQ ID NO: 28 is the determined cDNA sequence for 54182.1 which shares homology with Transforming growth factor-beta induced gene product.

30 SEQ ID NO: 29 is the determined cDNA sequence for 53989.1 which shares homology with GDP-mannose 4,6 dehydratase.

SEQ ID NO: 30 is the determined cDNA sequence for 54181.1.

SEQ ID NO: 31 is the determined cDNA sequence for 54079.1 which shares homology with PAC 75N13 on chromosome Xq21.1.

SEQ ID NO: 32 is the determined cDNA sequence for 54114.1 which shares homology with Mus fork head transcription factor gene.

5 SEQ ID NO: 33 is the determined cDNA sequence for 54160.1 which shares homology with Clone 146H21 on chromosome Xq22.

SEQ ID NO: 34 is the determined cDNA sequence for 54168.1 which shares homology with Glutamine:fructose-6-phosphate amidotransferase.

10 SEQ ID NO: 35 is the determined cDNA sequence for 54078.1 which shares homology with PAC 75N13 on chromosome Xq21.1.

SEQ ID NO: 36 is the determined cDNA sequence for 53900.1 which shares homology with Intestinal peptide-associated transporter HPT-1.

SEQ ID NO: 37 is the determined cDNA sequence for 54147.1.

15 SEQ ID NO: 38 is the determined cDNA sequence for 54033.1 which shares homology with Human proteinase activated receptor-2.

SEQ ID NO: 39 is the determined cDNA sequence for 53908.1 which shares homology with GalNAc-T3 gene.

SEQ ID NO: 40 is the determined cDNA sequence for 54022.1.

20 SEQ ID NO: 41 is the determined cDNA sequence for 54039.1 which shares homology with Constitutive fragile sequence.

SEQ ID NO: 42 is the determined cDNA sequence for 54037.1 which shares homology with CD24 signal transducer gene.

SEQ ID NO: 43 is the determined cDNA sequence for 54129.1 which shares homology with Human c-myb gene.

25 SEQ ID NO: 44 is the determined cDNA sequence for 54054.1 which shares homology with Pyrroline-t-carboxylate synthase long form.

SEQ ID NO: 45 is the determined cDNA sequence for 54055.1 which shares homology with Human zinc finger protein ZNF-139.

30 SEQ ID NO: 46 is the determined cDNA sequence for 54046.1 which shares homology with Gene for membrane cofactor protein.

SEQ ID NO: 47 is the determined cDNA sequence for 54047.1 which shares homology with Colon Kruppel-like factor.

SEQ ID NO: 48 is the determined cDNA sequence for 54040.1 which shares homology with Human capping protein alpha subunit isoform 1.

5 SEQ ID NO: 49 is the determined cDNA sequence for 54035.1 which shares homology with Ig lambda-chain.

SEQ ID NO: 50 is the determined cDNA sequence for 54130.1 which shares homology with Protein tyrosine kinase.

10 SEQ ID NO: 51 is the determined cDNA sequence for 54045.1 which shares homology with cDNA FLJ10610 fis, clone NT2RP2005293.

SEQ ID NO: 52 is the determined cDNA sequence for 54052.1 which shares homology with Human microtubule-associated protein 7.

SEQ ID NO: 53 is the determined cDNA sequence for 54050.1 which shares homology with Human retinoblastoma susceptibility protein.

15 SEQ ID NO: 54 is the determined cDNA sequence for 54051.1 which shares homology with Human reticulocalbin.

SEQ ID NO: 55 is the determined cDNA sequence for 54178.1 which shares homology with Translation initiation factor eIF3 p36 subunit.

20 SEQ ID NO: 56 is the determined cDNA sequence for 54148.1 which shares homology with Human apurinic/apyrimidinic-endonuclease.

SEQ ID NO: 57 is the determined cDNA sequence for 54058.1.

SEQ ID NO: 58 is the determined cDNA sequence for 54059.1 which shares homology with Human integral transmembrane protein 1.

25 SEQ ID NO: 59 is the determined cDNA sequence for 54126.1 which shares homology with Human serine kinase.

SEQ ID NO: 60 is the determined cDNA sequence for 54127.1 which shares homology with Human CG1-44 protein.

SEQ ID NO: 61 is the determined cDNA sequence for 54049.1 which shares homology with HADH/NADPH thyroid oxidase p138-tox protein.

30 SEQ ID NO: 62 is the determined cDNA sequence for 54056.1 which shares homology with Human peptide transporter (TAP1) protein.

SEQ ID NO: 63 is the determined cDNA sequence for 54064.1 which shares homology with Clone RP1-39G22 on chromosome 1p32.1-34.3.

SEQ ID NO: 64 is the determined cDNA sequence for 54124.1 which shares homology with Clone Transforming growth factor-beta induced gene product.

5 SEQ ID NO: 65 is the determined cDNA sequence for 54063.1.

SEQ ID NO: 66 is the determined cDNA sequence for 54141.1 which shares homology with Cytokeratin 8.

SEQ ID NO: 67 is the determined cDNA sequence for 54119.1 which shares homology with Human coat protein gamma-cop.

10 SEQ ID NO: 68 is the determined cDNA sequence for 54111.1 which shares homology with Bumetanide-sensitive Na-K-Cl cotransporter.

SEQ ID NO: 69 is the determined cDNA sequence for 54121.1 which shares homology with cDNA FLJ10969 fis, clone PLACE1000909.

15 SEQ ID NO: 70 is the determined cDNA sequence for 54065.1 which shares homology with BAC clone 215O12.

SEQ ID NO: 71 is the determined cDNA sequence for 54060.1 which shares homology with Autoantigen calreticulin.

SEQ ID NO: 72 is the determined cDNA sequence for 54125.1 which shares homology with Human hepatic squalene synthetase.

20 SEQ ID NO: 73 is the determined cDNA sequence for 54143.1 which shares homology with Human RAD21 homolog.

SEQ ID NO: 74 is the determined cDNA sequence for 54139.1 which shares homology with Human MHC class II HLA-DR-alpha.

25 SEQ ID NO: 75 is the determined cDNA sequence for 54137.1 which shares homology with Human Claudin-7.

SEQ ID NO: 76 is the determined cDNA sequence for 54044.1 which shares homology with Ribosome protein S6 kinase 1.

SEQ ID NO: 77 is the determined cDNA sequence for 54042.1 which shares homology with CO-029 tumor associated antigen.

30 SEQ ID NO: 78 is the determined cDNA sequence for 54043.1 which shares homology with KIAA1077 protein.

SEQ ID NO: 79 is the determined cDNA sequence for 54136.1 which shares homology with Human lipocortin II.

SEQ ID NO: 80 is the determined cDNA sequence for 54157.1 which shares homology with PAC 454G6 on chromosome 1q24.

5 SEQ ID NO: 81 is the determined cDNA sequence for 54140.1.

SEQ ID NO: 82 is the determined cDNA sequence for 54120.1.

SEQ ID NO: 83 is the determined cDNA sequence for 54145.1 which shares homology with KIAA0152.

10 SEQ ID NO: 84 is the determined cDNA sequence for 54117.1 which shares homology with Tumor antigen L6.

SEQ ID NO: 85 is the determined cDNA sequence for 54116.1 which shares homology with UDP-N-acetylglucosamine transporter.

SEQ ID NO: 86 is the determined cDNA sequence for 54151.1.

15 SEQ ID NO: 87 is the determined cDNA sequence for 54152.1 which shares homology with Cystine/glutamate transporter.

SEQ ID NO: 88 is the determined cDNA sequence for 54115.1.

SEQ ID NO: 89 is the determined cDNA sequence for 54146.1 which shares homology with GAPDH.

20 SEQ ID NO: 90 is the determined cDNA sequence for 54155.1 which shares homology with cDNA DKFZp586O0118.

SEQ ID NO: 91 is the determined cDNA sequence for 54159.1.

SEQ ID NO: 92 is the determined cDNA sequence for 54020.1 which shares homology with Neutrophil lipocalin.

25 SEQ ID NO: 93 is the determined cDNA sequence for 54169.1 which shares homology with Nuclear matrix protein NRP/B.

SEQ ID NO: 94 is the determined cDNA sequence for 54167.1 which shares homology with CGI-151/KIAA0992 protein.

SEQ ID NO: 95 is the determined cDNA sequence for 54030.1.

SEQ ID NO: 96 is the determined cDNA sequence for 54161.1.

30 SEQ ID NO: 97 is the determined cDNA sequence for 54162.1 which shares homology with Poly A binding protein.

SEQ ID NO: 98 is the determined cDNA sequence for 54163.1 which shares homology with Ribosome protein L13.

SEQ ID NO: 99 is the determined cDNA sequence for 54164.1 which shares homology with Human alpha enolase.

5 SEQ ID NO: 100 is the determined cDNA sequence for 54132.1 which shares homology with Human E-1 enzyme.

SEQ ID NO: 101 is the determined cDNA sequence for 54112.1 which shares homology with cDNA DKFZp58612022.

10 SEQ ID NO: 102 is the determined cDNA sequence for 54133.1 which shares homology with Human ZW10 interactor Zwint.

SEQ ID NO: 103 is the determined cDNA sequence for 54165.1 which shares homology with Bumetanide-sensitive Na-K-Cl cotransporter.

SEQ ID NO: 104 is the determined cDNA sequence for 54158.1 which shares homology with cDNA FLJ10549 fis, clone NT2RP2001976.

15 SEQ ID NO: 105 is the determined cDNA sequence for 54131.1 which shares homology with cDNA DKFZp434C0523.

SEQ ID NO: 106 is the determined cDNA sequence for 54122.1.

SEQ ID NO: 107 is the determined cDNA sequence for 54098.1.

20 SEQ ID NO: 108 is the determined cDNA sequence for 54173.1 which shares homology with NADH-ubiquinone oxidoreductase NDUFS2 subunit.

SEQ ID NO: 109 is the determined cDNA sequence for 54108.1 which shares homology with Phospholipase A2.

SEQ ID NO: 110 is the determined cDNA sequence for 54175.1 which shares homology with cDNA FLJ10610 fis, clone NT2RP2005293.

25 SEQ ID NO: 111 is the determined cDNA sequence for 54179.1 which shares homology with Ig heavy chain variable region.

SEQ ID NO: 112 is the determined cDNA sequence for 54177.1 which shares homology with Protein phosphatase 2C gamma.

30 SEQ ID NO: 113 is the determined cDNA sequence for 54170.1 which shares homology with Cyclin protein.

SEQ ID NO: 114 is the determined cDNA sequence for 54176.1 which shares homology with Transgelin 2 (predicted).

SEQ ID NO: 115 is the determined cDNA sequence for 54180.1 which shares homology with Human GalNAc-T3 gene.

5 SEQ ID NO: 116 is the determined cDNA sequence for 53897.1 which shares homology with cDNA FLJ10884 fis, clone NT2RP4001950.

SEQ ID NO: 117 is the determined cDNA sequence for 54027.1.

SEQ ID NO: 118 is the determined cDNA sequence for 54183.1 which shares homology with Alpha topoisomerase truncated form.

10 SEQ ID NO: 119 is the determined cDNA sequence for 54107.1 which shares homology with KIAA 1289.

SEQ ID NO: 120 is the determined cDNA sequence for 54106.1 which shares homology with AD022 protein.

SEQ ID NO: 121 is the determined cDNA sequence for 53902.1.

15 SEQ ID NO: 122 is the determined cDNA sequence for 53918.1 which shares homology with Chromosome 17, clone hRPK.692_E_18.

SEQ ID NO: 123 is the determined cDNA sequence for 53904.1.

SEQ ID NO: 124 is the determined cDNA sequence for 53910.1 which shares homology with cDNA FLJ10823 fis, clone NT2RP4001080.

20 SEQ ID NO: 125 is the determined cDNA sequence for 53903.1 which shares homology with Vector.

SEQ ID NO: 126 is the determined cDNA sequence for 54103.1.

SEQ ID NO: 127 is the determined cDNA sequence for 53917.1 which shares homology with Cytochrome P450 IIIA4.

25 SEQ ID NO: 128 is the determined cDNA sequence for 54004.1 which shares homology with CEA.

SEQ ID NO: 129 is the determined cDNA sequence for 53913.1 which shares homology with Protein phosphatase (KAP1).

SEQ ID NO: 130 is the determined cDNA sequence for 54134.1.

30 SEQ ID NO: 131 is the determined cDNA sequence for 53999.1 which shares homology with Alpha enolase.

SEQ ID NO: 132 is the determined cDNA sequence for 53938.1 which shares homology with Histone deacetylase HD1.

SEQ ID NO: 133 is the determined cDNA sequence for 53939.1 which shares homology with citb_338_f_24, complete sequence.

5 SEQ ID NO: 134 is the determined cDNA sequence for 53928.1 which shares homology with Human squalene epoxidase.

SEQ ID NO: 135 is the determined cDNA sequence for 53914.1 which shares homology with Human aspartyl-tRNA-synthetase alpha-2 subunit.

10 SEQ ID NO: 136 is the determined cDNA sequence for 53915.1 which shares homology with Gamma-actin.

SEQ ID NO: 137 is the determined cDNA sequence for 54101.1 which shares homology with Human AP-mu chain family member mu1B.

SEQ ID NO: 138 is the determined cDNA sequence for 53922.1 which shares homology with Human Cctg mRNA for chaperonin.

15 SEQ ID NO: 139 is the determined cDNA sequence for 54023.1 which shares homology with Chromosome 19.

SEQ ID NO: 140 is the determined cDNA sequence for 53930.1 which shares homology with Human MEGF7.

20 SEQ ID NO: 141 is the determined cDNA sequence for 53921.1 which shares homology with Connexin 26.

SEQ ID NO: 142 is the determined cDNA sequence for 54002.1 which shares homology with Human dipeptidyl peptidase IV.

SEQ ID NO: 143 is the determined cDNA sequence for 54003.1 which shares homology with Chromosome 5 clone CTC-436P18.

25 SEQ ID NO: 144 is the determined cDNA sequence for 54005.1 which shares homology with Human 2-oxoglutarate dehydrogenase.

SEQ ID NO: 145 is the determined cDNA sequence for 53925.1 which shares homology with RHO guanine nucleotide-exchange factor.

30 SEQ ID NO: 146 is the determined cDNA sequence for 53927.1 which shares homology with 12q24 PAC RPC11-261P5.

SEQ ID NO: 147 is the determined cDNA sequence for 54083.1 which shares homology with Human colon mucosa-associated mRNA.

SEQ ID NO: 148 is the determined cDNA sequence for 53937.1.

5 SEQ ID NO: 149 is the determined cDNA sequence for 54074.1 which shares homology with Clone RP4-621F18 on chromosome 1p11.4-21.3.

SEQ ID NO: 150 is the determined cDNA sequence for 54105.1.

SEQ ID NO: 151 is the determined cDNA sequence for 53961.1 which shares homology with Human embryonic lung protein.

SEQ ID NO: 152 is the determined cDNA sequence for 53919.1.

10 SEQ ID NO: 153 is the determined cDNA sequence for 53933.1 which shares homology with Human leukocyte surface protein CD31.

SEQ ID NO: 154 is the determined cDNA sequence for 53972.1 which shares homology with cDNA FLJ10679 fis, clone NT2RP2006565.

SEQ ID NO: 155 is the determined cDNA sequence for 53906.1.

15 SEQ ID NO: 156 is the determined cDNA sequence for 53924.1 which shares homology with Poly A binding protein.

SEQ ID NO: 157 is the determined cDNA sequence for 54144.1.

SEQ ID NO: 158 is the determined cDNA sequence for 54068.1 which shares homology with Cystic fibrosis transmembrane conductance regulator.

20 SEQ ID NO: 159 is the determined cDNA sequence for 53929.1.

SEQ ID NO: 160 is the determined cDNA sequence for 53959.1 which shares homology with KIAA1050.

SEQ ID NO: 161 is the determined cDNA sequence for 53942.1.

25 SEQ ID NO: 162 is the determined cDNA sequence for 53931.1 which shares homology with cDNA FLJ11127 fis, clone PLACE 1006225.

SEQ ID NO: 163 is the determined cDNA sequence for 53935.1 which shares homology with Human set gene.

SEQ ID NO: 164 is the determined cDNA sequence for 54099.1 which shares homology with Human pleckstrin 2.

30 SEQ ID NO: 165 is the determined cDNA sequence for 53943.1 which shares homology with KIAA0965.

SEQ ID NO: 166 is the determined cDNA sequence for 54000.1 which shares homology with Tis 11d gene.

SEQ ID NO: 167 is the determined cDNA sequence for 54100.1 which shares homology with Cyhtokine (GRO-gamma).

5 SEQ ID NO: 168 is the determined cDNA sequence for 53940.1 which shares homology with Human p85Mcm mRNA.

SEQ ID NO: 169 is the determined cDNA sequence for 53941.1 which shares homology with cDNA DKFZp586H0519.

10 SEQ ID NO: 170 is the determined cDNA sequence for 53953.1 which shares homology with SOX9.

SEQ ID NO: 171 is the determined cDNA sequence for 54007.1 which shares homology with VAV-like protein.

SEQ ID NO: 172 is the determined cDNA sequence for 53950.1 which shares homology with NF-E2 related factor 3.

15 SEQ ID NO: 173 is the determined cDNA sequence for 53968.1 which shares homology with cDNA FLJ20127 fis, clone COL06176.

SEQ ID NO: 174 is the determined cDNA sequence for 53945.1.

SEQ ID NO: 175 is the determined cDNA sequence for 54091.1.

20 SEQ ID NO: 176 is the determined cDNA sequence for 54013.1 which shares homology with Human argininosuccinate synthetase.

SEQ ID NO: 177 is the determined cDNA sequence for 54092.1 which shares homology with Human serine kinase.

SEQ ID NO: 178 is the determined cDNA sequence for 54095.1 which shares homology with Clone RP1-155G6 on chromosome 20.

25 SEQ ID NO: 179 is the determined cDNA sequence for 53987.1 which shares homology with Human phospholipase C beta 4.

SEQ ID NO: 180 is the determined cDNA sequence for 53967.1.

SEQ ID NO: 181 is the determined cDNA sequence for 53963.1 which shares homology with VAV-3 protein.

30 SEQ ID NO: 182 is the determined cDNA sequence for 54032.1.

SEQ ID NO: 183 is the determined cDNA sequence for 54067.1 which shares homology with PAC RPCI-1 133G21 map 21q11.1 region D21S190.

SEQ ID NO: 184 is the determined cDNA sequence for 54057.1 which shares homology with Calcium-binding protein S100P.

5 SEQ ID NO: 185 is the determined cDNA sequence for 54135.1 which shares homology with Human leupaxin.

SEQ ID NO: 186 is the determined cDNA sequence for 53969.1 which shares homology with VAV-3 Protein.

SEQ ID NO: 187 is the determined cDNA sequence for 53970.1.

10 SEQ ID NO: 188 is the determined cDNA sequence for 53966.1 which shares homology with hnRNP type A/B protein.

SEQ ID NO: 189 is the determined cDNA sequence for 53995.1 which shares homology with Human cell cycle control gene CDC2.

SEQ ID NO: 190 is the determined cDNA sequence for 54075.1.

15 SEQ ID NO: 191 is the determined cDNA sequence for 54094.1.

SEQ ID NO: 192 is the determined cDNA sequence for 53977.1.

SEQ ID NO: 193 is the determined cDNA sequence for 54123.1 which shares homology with BAC clone RG083M05 from 7q21-7q22.

20 SEQ ID NO: 194 is the determined cDNA sequence for 53960.1 which shares homology with Human STS WI-14644.

SEQ ID NO: 195 is the determined cDNA sequence for 53976.1 which shares homology with Human glutaminyl-tRNA synthetase.

SEQ ID NO: 196 is the determined cDNA sequence for 54096.1 which shares homology with Human 26S proteasome-associated pad 1 homolog.

25 SEQ ID NO: 197 is the determined cDNA sequence for 54110.1 which shares homology with Human squalene epoxidase.

SEQ ID NO: 198 is the determined cDNA sequence for 53920.1 which shares homology with Human nuclear chloride ion channel protein.

30 SEQ ID NO: 199 is the determined cDNA sequence for 53979.1 which shares homology with PAC RPCI-1 133G21 map 21q11.1 region D21S190.

SEQ ID NO: 200 is the determined cDNA sequence for 54081.1 which shares homology with PAC clone RP5-1185I7 from 7q11.23-q21.

SEQ ID NO: 201 is the determined cDNA sequence for 54082.1 which shares homology with Human ephrin.

5 SEQ ID NO: 202 is the determined cDNA sequence for 53986.1 which shares homology with cDNA FLJ20673 fis, clone KAIA4464.

SEQ ID NO: 203 is the determined cDNA sequence for 53992.1.

SEQ ID NO: 204 is the determined cDNA sequence for 54016.1.

10 SEQ ID NO: 205 is the determined cDNA sequence for 54018.1 which shares homology with CD9 antigen.

SEQ ID NO: 206 is the determined cDNA sequence for 53985.1 which shares homology with KIAA0715.

SEQ ID NO: 207 is the determined cDNA sequence for 53973.1 which shares homology with Cyclin B.

15 SEQ ID NO: 208 is the determined cDNA sequence for 54012.1 which shares homology with KIAA1225.

SEQ ID NO: 209 is the determined cDNA sequence for 53982.1.

SEQ ID NO: 210 is the determined cDNA sequence for 53988.1 which shares homology with Colon mucosa-associated mRNA.

20 SEQ ID NO: 211 is the determined cDNA sequence for 53990.1 which shares homology with cDNA FLJ20171 fis, clone COL09761.

SEQ ID NO: 212 is the determined cDNA sequence for 53991.1.

SEQ ID NO: 213 is the determined cDNA sequence for 51519.1 which shares homology with CEA.

25 SEQ ID NO: 214 is the determined cDNA sequence for 51507.1 which shares homology with Adenocarcinoma-associated antigen.

SEQ ID NO: 215 is the determined cDNA sequence for 51435.1 which shares homology with Secreted protein XAG.

30 SEQ ID NO: 216 is the determined cDNA sequence for 51425.1 which shares homology with Adenocarcinoma-associated antigen.

SEQ ID NO: 217 is the determined cDNA sequence for 51548.1.

SEQ ID NO: 218 is the determined cDNA sequence for 51430.1 which shares homology with CEA.

SEQ ID NO: 219 is the determined cDNA sequence for 51549.1 which shares homology with CEA.

5 SEQ ID NO: 220 is the determined cDNA sequence for 51439.1 which shares homology with Nonspecific crossreacting antigen.

SEQ ID NO: 221 is the determined cDNA sequence for 51535.1 which shares homology with Neutrophil gelatinase associated lipocalin.

10 SEQ ID NO: 222 is the determined cDNA sequence for 51486.1 which shares homology with Transformation growth factor-beta induced gene product.

SEQ ID NO: 223 is the determined cDNA sequence for 51479.1 which shares homology with Undetermined origin found 5' to NCA mRNA.

SEQ ID NO: 224 is the determined cDNA sequence for 51469.1 which shares homology with Galectin-4.

15 SEQ ID NO: 225 is the determined cDNA sequence for 51470.1 which shares homology with Nonspecific crossreacting antigen.

SEQ ID NO: 226 is the determined cDNA sequence for 51536.1 which shares homology with Secreted protein XAG.

20 SEQ ID NO: 227 is the determined cDNA sequence for 51483.1 which shares homology with Clone 146H21 on chromosome Xq22.

SEQ ID NO: 228 is the determined cDNA sequence for 51522.1 which shares homology with GAPDH.

SEQ ID NO: 229 is the determined cDNA sequence for 51485.1 which shares homology with Mucin 11.

25 SEQ ID NO: 230 is the determined cDNA sequence for 51460.1 which shares homology with Nonspecific crossreacting antigen.

SEQ ID NO: 231 is the determined cDNA sequence for 51458.1 which shares homology with KIAA0517 protein.

30 SEQ ID NO: 232 is the determined cDNA sequence for 51506.1 which shares homology with Surface glycoprotein CD44.

SEQ ID NO: 233 is the determined cDNA sequence for 51440.1 which shares homology with Chromosome 21q22.1, D21S226-AML region.

SEQ ID NO: 234 is the determined cDNA sequence for C907P.

SEQ ID NO: 235 is the amino acid sequence for C907P.

5 SEQ ID NO: 236 is the determine cDNA sequence for Ra12-C915P-f3.

SEQ ID NO: 237 is the amino acid sequence for Ra12-C915P-f3.

SEQ ID NO: 238 is the nucleotide sequence of the AW154 primer.

SEQ ID NO: 239 is the nucleotide sequence of the AW155 primer.

SEQ ID NO: 240 is the nucleotide sequence of the AW156 primer.

10 SEQ ID NO: 241 is the nucleotide sequence of the AW157 primer.

SEQ ID NO: 242 is the nucleotide sequence of the AW158 primer.

SEQ ID NO: 243 is the nucleotide sequence of the AW159 primer.

SEQ ID NO: 244 is the determined full-length cDNA sequence of C915P.

15 SEQ ID NO: 245 is the amino acid sequence encoded by the cDNA sequence set forth in SEQ ID NO:244.

DETAILED DESCRIPTION OF THE INVENTION

20 The present invention is directed generally to compositions and their use in the therapy and diagnosis of cancer, particularly colon cancer. As described further below, illustrative compositions of the present invention include, but are not restricted to, polypeptides, particularly immunogenic polypeptides, polynucleotides encoding such polypeptides, antibodies and other binding agents, antigen presenting cells (APCs) and immune system cells (*e.g.*, T cells).

25 The practice of the present invention will employ, unless indicated specifically to the contrary, conventional methods of virology, immunology, microbiology, molecular biology and recombinant DNA techniques within the skill of the art, many of which are described below for the purpose of illustration. Such techniques are explained fully in the literature. See, *e.g.*, Sambrook, et al. Molecular Cloning: A Laboratory Manual (2nd Edition, 1989); Maniatis et al. Molecular Cloning: A Laboratory Manual (1982); DNA Cloning: A Practical Approach, vol. I & II (D.

30

Glover, ed.); Oligonucleotide Synthesis (N. Gait, ed., 1984); Nucleic Acid Hybridization (B. Hames & S. Higgins, eds., 1985); Transcription and Translation (B. Hames & S. Higgins, eds., 1984); Animal Cell Culture (R. Freshney, ed., 1986); Perbal, A Practical Guide to Molecular Cloning (1984).

5 All publications, patents and patent applications cited herein, whether supra or infra, are hereby incorporated by reference in their entirety.

As used in this specification and the appended claims, the singular forms "a," "an" and "the" include plural references unless the content clearly dictates otherwise.

10 POLYPEPTIDE COMPOSITIONS

As used herein, the term "polypeptide" is used in its conventional meaning, *i.e.*, as a sequence of amino acids. The polypeptides are not limited to a specific length of the product; thus, peptides, oligopeptides, and proteins are included within the definition of polypeptide, and such terms may be used interchangeably
15 herein unless specifically indicated otherwise. This term also does not refer to or exclude post-expression modifications of the polypeptide, for example, glycosylations, acetylations, phosphorylations and the like, as well as other modifications known in the art, both naturally occurring and non-naturally occurring. A polypeptide may be an entire protein, or a subsequence thereof. Particular polypeptides of interest in the
20 context of this invention are amino acid subsequences comprising epitopes, *i.e.*, antigenic determinants substantially responsible for the immunogenic properties of a polypeptide and being capable of evoking an immune response.

Particularly illustrative polypeptides of the present invention comprise those encoded by a polynucleotide sequence set forth in any one of SEQ ID NOs:1-234,
25 236, and 244, or a sequence that hybridizes under moderately stringent conditions, or, alternatively, under highly stringent conditions, to a polynucleotide sequence set forth in any one of SEQ ID NOs:1-234, 236, and 244. Certain other illustrative polypeptides of the invention comprise amino acid sequences as set forth in any one of SEQ ID NOs:235, 237, and 245.

The polypeptides of the present invention are sometimes herein referred to as colon tumor proteins or colon tumor polypeptides, as an indication that their identification has been based at least in part upon their increased levels of expression in colon tumor samples. Thus, a "colon tumor polypeptide" or "colon tumor protein," refers generally to a polypeptide sequence of the present invention, or a polynucleotide sequence encoding such a polypeptide, that is expressed in a substantial proportion of colon tumor samples, for example preferably greater than about 20%, more preferably greater than about 30%, and most preferably greater than about 50% or more of colon tumor samples tested, at a level that is at least two fold, and preferably at least five fold, greater than the level of expression in normal tissues, as determined using a representative assay provided herein. A colon tumor polypeptide sequence of the invention, based upon its increased level of expression in tumor cells, has particular utility both as a diagnostic marker as well as a therapeutic target, as further described below.

In certain preferred embodiments, the polypeptides of the invention are immunogenic, *i.e.*, they react detectably within an immunoassay (such as an ELISA or T-cell stimulation assay) with antisera and/or T-cells from a patient with colon cancer. Screening for immunogenic activity can be performed using techniques well known to the skilled artisan. For example, such screens can be performed using methods such as those described in Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988. In one illustrative example, a polypeptide may be immobilized on a solid support and contacted with patient sera to allow binding of antibodies within the sera to the immobilized polypeptide. Unbound sera may then be removed and bound antibodies detected using, for example, ¹²⁵I-labeled Protein A.

As would be recognized by the skilled artisan, immunogenic portions of the polypeptides disclosed herein are also encompassed by the present invention. An "immunogenic portion," as used herein, is a fragment of an immunogenic polypeptide of the invention that itself is immunologically reactive (*i.e.*, specifically binds) with the B-cells and/or T-cell surface antigen receptors that recognize the polypeptide. Immunogenic portions may generally be identified using well known techniques, such as those summarized in Paul, *Fundamental Immunology*, 3rd ed., 243-247 (Raven Press,

1993) and references cited therein. Such techniques include screening polypeptides for the ability to react with antigen-specific antibodies, antisera and/or T-cell lines or clones. As used herein, antisera and antibodies are "antigen-specific" if they specifically bind to an antigen (*i.e.*, they react with the protein in an ELISA or other immunoassay, and do not react detectably with unrelated proteins). Such antisera and antibodies may be prepared as described herein, and using well-known techniques.

In one preferred embodiment, an immunogenic portion of a polypeptide of the present invention is a portion that reacts with antisera and/or T-cells at a level that is not substantially less than the reactivity of the full-length polypeptide (*e.g.*, in an ELISA and/or T-cell reactivity assay). Preferably, the level of immunogenic activity of the immunogenic portion is at least about 50%, preferably at least about 70% and most preferably greater than about 90% of the immunogenicity for the full-length polypeptide. In some instances, preferred immunogenic portions will be identified that have a level of immunogenic activity greater than that of the corresponding full-length polypeptide, *e.g.*, having greater than about 100% or 150% or more immunogenic activity.

In certain other embodiments, illustrative immunogenic portions may include peptides in which an N-terminal leader sequence and/or transmembrane domain have been deleted. Other illustrative immunogenic portions will contain a small N- and/or C-terminal deletion (*e.g.*, 1-30 amino acids, preferably 5-15 amino acids), relative to the mature protein.

In another embodiment, a polypeptide composition of the invention may also comprise one or more polypeptides that are immunologically reactive with T cells and/or antibodies generated against a polypeptide of the invention, particularly a polypeptide having an amino acid sequence disclosed herein, or to an immunogenic fragment or variant thereof.

In another embodiment of the invention, polypeptides are provided that comprise one or more polypeptides that are capable of eliciting T cells and/or antibodies that are immunologically reactive with one or more polypeptides described herein, or one or more polypeptides encoded by contiguous nucleic acid sequences contained in the polynucleotide sequences disclosed herein, or immunogenic fragments

or variants thereof, or to one or more nucleic acid sequences which hybridize to one or more of these sequences under conditions of moderate to high stringency.

The present invention, in another aspect, provides polypeptide fragments comprising at least about 5, 10, 15, 20, 25, 50, or 100 contiguous amino acids, or more, including all intermediate lengths, of a polypeptide compositions set forth herein, such as those set forth in SEQ ID NOs:235, 237, and 245, or those encoded by a polynucleotide sequence set forth in a sequence of SEQ ID NOs:1-234, 236, and 244.

In another aspect, the present invention provides variants of the polypeptide compositions described herein. Polypeptide variants generally encompassed by the present invention will typically exhibit at least about 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% or more identity (determined as described below), along its length, to a polypeptide sequences set forth herein.

In one preferred embodiment, the polypeptide fragments and variants provided by the present invention are immunologically reactive with an antibody and/or T-cell that reacts with a full-length polypeptide specifically set forth herein.

In another preferred embodiment, the polypeptide fragments and variants provided by the present invention exhibit a level of immunogenic activity of at least about 50%, preferably at least about 70%, and most preferably at least about 90% or more of that exhibited by a full-length polypeptide sequence specifically set forth herein.

A polypeptide "variant," as the term is used herein, is a polypeptide that typically differs from a polypeptide specifically disclosed herein in one or more substitutions, deletions, additions and/or insertions. Such variants may be naturally occurring or may be synthetically generated, for example, by modifying one or more of the above polypeptide sequences of the invention and evaluating their immunogenic activity as described herein and/or using any of a number of techniques well known in the art.

For example, certain illustrative variants of the polypeptides of the invention include those in which one or more portions, such as an N-terminal leader sequence or transmembrane domain, have been removed. Other illustrative variants

include variants in which a small portion (*e.g.*, 1-30 amino acids, preferably 5-15 amino acids) has been removed from the N- and/or C-terminal of the mature protein.

In many instances, a variant will contain conservative substitutions. A "conservative substitution" is one in which an amino acid is substituted for another amino acid that has similar properties, such that one skilled in the art of peptide chemistry would expect the secondary structure and hydrophobic nature of the polypeptide to be substantially unchanged. As described above, modifications may be made in the structure of the polynucleotides and polypeptides of the present invention and still obtain a functional molecule that encodes a variant or derivative polypeptide with desirable characteristics, *e.g.*, with immunogenic characteristics. When it is desired to alter the amino acid sequence of a polypeptide to create an equivalent, or even an improved, immunogenic variant or portion of a polypeptide of the invention, one skilled in the art will typically change one or more of the codons of the encoding DNA sequence according to Table 1.

For example, certain amino acids may be substituted for other amino acids in a protein structure without appreciable loss of interactive binding capacity with structures such as, for example, antigen-binding regions of antibodies or binding sites on substrate molecules. Since it is the interactive capacity and nature of a protein that defines that protein's biological functional activity, certain amino acid sequence substitutions can be made in a protein sequence, and, of course, its underlying DNA coding sequence, and nevertheless obtain a protein with like properties. It is thus contemplated that various changes may be made in the peptide sequences of the disclosed compositions, or corresponding DNA sequences which encode said peptides without appreciable loss of their biological utility or activity.

25

TABLE 1

| Amino Acids | | | Codons | | | | | | |
|---------------|-----|---|--------|-----|-----|-----|-----|-----|--|
| Alanine | Ala | A | GCA | GCC | GCG | GCU | | | |
| Cysteine | Cys | C | UGC | UGU | | | | | |
| Aspartic acid | Asp | D | GAC | GAU | | | | | |
| Glutamic acid | Glu | E | GAA | GAG | | | | | |
| Phenylalanine | Phe | F | UUC | UUU | | | | | |
| Glycine | Gly | G | GGA | GGC | GGG | GGU | | | |
| Histidine | His | H | CAC | CAU | | | | | |
| Isoleucine | Ile | I | AUA | AUC | AUU | | | | |
| Lysine | Lys | K | AAA | AAG | | | | | |
| Leucine | Leu | L | UUA | UUG | CUA | CUC | CUG | CUU | |
| Methionine | Met | M | AUG | | | | | | |
| Asparagine | Asn | N | AAC | AAU | | | | | |
| Proline | Pro | P | CCA | CCC | CCG | CCU | | | |
| Glutamine | Gln | Q | CAA | CAG | | | | | |
| Arginine | Arg | R | AGA | AGG | CGA | CGC | CGG | CGU | |
| Serine | Ser | S | AGC | AGU | UCA | UCC | UCG | UCU | |
| Threonine | Thr | T | ACA | ACC | ACG | ACU | | | |
| Valine | Val | V | GUA | GUC | GUG | GUU | | | |
| Tryptophan | Trp | W | UGG | | | | | | |
| Tyrosine | Tyr | Y | UAC | UAU | | | | | |

In making such changes, the hydropathic index of amino acids may be considered. The importance of the hydropathic amino acid index in conferring interactive biologic function on a protein is generally understood in the art (Kyte and Doolittle, 1982, incorporated herein by reference). It is accepted that the relative hydropathic character of the amino acid contributes to the secondary structure of the resultant protein, which in turn defines the interaction of the protein with other molecules, for example, enzymes, substrates, receptors, DNA, antibodies, antigens, and the like. Each amino acid has been assigned a hydropathic index on the basis of its hydrophobicity and charge characteristics (Kyte and Doolittle, 1982). These values are:

isoleucine (+4.5); valine (+4.2); leucine (+3.8); phenylalanine (+2.8); cysteine/cystine (+2.5); methionine (+1.9); alanine (+1.8); glycine (−0.4); threonine (−0.7); serine (−0.8); tryptophan (−0.9); tyrosine (−1.3); proline (−1.6); histidine (−3.2); glutamate (−3.5); glutamine (−3.5); aspartate (−3.5); asparagine (−3.5); lysine (−3.9); and arginine (−4.5).

It is known in the art that certain amino acids may be substituted by other amino acids having a similar hydropathic index or score and still result in a protein with similar biological activity, *i.e.* still obtain a biological functionally equivalent protein. In making such changes, the substitution of amino acids whose hydropathic indices are within ± 2 is preferred, those within ± 1 are particularly preferred, and those within ± 0.5 are even more particularly preferred. It is also understood in the art that the substitution of like amino acids can be made effectively on the basis of hydrophilicity. U. S. Patent 4,554,101 (specifically incorporated herein by reference in its entirety), states that the greatest local average hydrophilicity of a protein, as governed by the hydrophilicity of its adjacent amino acids, correlates with a biological property of the protein.

As detailed in U. S. Patent 4,554,101, the following hydrophilicity values have been assigned to amino acid residues: arginine (+3.0); lysine (+3.0); aspartate (+3.0 \pm 1); glutamate (+3.0 \pm 1); serine (+0.3); asparagine (+0.2); glutamine (+0.2); glycine (0); threonine (−0.4); proline (−0.5 \pm 1); alanine (−0.5); histidine (−0.5); cysteine (−1.0); methionine (−1.3); valine (−1.5); leucine (−1.8); isoleucine (−1.8); tyrosine (−2.3); phenylalanine (−2.5); tryptophan (−3.4). It is understood that an amino acid can be substituted for another having a similar hydrophilicity value and still obtain a biologically equivalent, and in particular, an immunologically equivalent protein. In such changes, the substitution of amino acids whose hydrophilicity values are within ± 2 is preferred, those within ± 1 are particularly preferred, and those within ± 0.5 are even more particularly preferred.

As outlined above, amino acid substitutions are generally therefore based on the relative similarity of the amino acid side-chain substituents, for example, their hydrophobicity, hydrophilicity, charge, size, and the like. Exemplary substitutions that take various of the foregoing characteristics into consideration are well known to those

of skill in the art and include: arginine and lysine; glutamate and aspartate; serine and threonine; glutamine and asparagine; and valine, leucine and isoleucine.

In addition, any polynucleotide may be further modified to increase stability *in vivo*. Possible modifications include, but are not limited to, the addition of
5 flanking sequences at the 5' and/or 3' ends; the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages in the backbone; and/or the inclusion of nontraditional bases such as inosine, queosine and wybutosine, as well as acetyl-methyl-, thio- and other modified forms of adenine, cytidine, guanine, thymine and uridine.

10 Amino acid substitutions may further be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity and/or the amphipathic nature of the residues. For example, negatively charged amino acids include aspartic acid and glutamic acid; positively charged amino acids include lysine and arginine; and amino acids with uncharged polar head groups having similar hydrophilicity values
15 include leucine, isoleucine and valine; glycine and alanine; asparagine and glutamine; and serine, threonine, phenylalanine and tyrosine. Other groups of amino acids that may represent conservative changes include: (1) ala, pro, gly, glu, asp, gln, asn, ser, thr; (2) cys, ser, tyr, thr; (3) val, ile, leu, met, ala, phe; (4) lys, arg, his; and (5) phe, tyr, trp, his. A variant may also, or alternatively, contain nonconservative changes. In a
20 preferred embodiment, variant polypeptides differ from a native sequence by substitution, deletion or addition of five amino acids or fewer. Variants may also (or alternatively) be modified by, for example, the deletion or addition of amino acids that have minimal influence on the immunogenicity, secondary structure and hydrophobic nature of the polypeptide.

25 As noted above, polypeptides may comprise a signal (or leader) sequence at the N-terminal end of the protein, which co-translationally or post-translationally directs transfer of the protein. The polypeptide may also be conjugated to a linker or other sequence for ease of synthesis, purification or identification of the polypeptide (e.g., poly-His), or to enhance binding of the polypeptide to a solid support.
30 For example, a polypeptide may be conjugated to an immunoglobulin Fc region.

When comparing polypeptide sequences, two sequences are said to be "identical" if the sequence of amino acids in the two sequences is the same when aligned for maximum correspondence, as described below. Comparisons between two sequences are typically performed by comparing the sequences over a comparison window to identify and compare local regions of sequence similarity. A "comparison window" as used herein, refers to a segment of at least about 20 contiguous positions, usually 30 to about 75, 40 to about 50, in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned.

Optimal alignment of sequences for comparison may be conducted using the Megalign program in the Lasergene suite of bioinformatics software (DNASTAR, Inc., Madison, WI), using default parameters. This program embodies several alignment schemes described in the following references: Dayhoff, M.O. (1978) A model of evolutionary change in proteins – Matrices for detecting distant relationships. In Dayhoff, M.O. (ed.) Atlas of Protein Sequence and Structure, National Biomedical Research Foundation, Washington DC Vol. 5, Suppl. 3, pp. 345-358; Hein J. (1990) Unified Approach to Alignment and Phylogenies pp. 626-645 *Methods in Enzymology* vol. 183, Academic Press, Inc., San Diego, CA; Higgins, D.G. and Sharp, P.M. (1989) *CABIOS* 5:151-153; Myers, E.W. and Muller W. (1988) *CABIOS* 4:11-17; Robinson, E.D. (1971) *Comb. Theor* 11:105; Saitou, N. Nei, M. (1987) *Mol. Biol. Evol.* 4:406-425; Sneath, P.H.A. and Sokal, R.R. (1973) *Numerical Taxonomy – the Principles and Practice of Numerical Taxonomy*, Freeman Press, San Francisco, CA; Wilbur, W.J. and Lipman, D.J. (1983) *Proc. Natl. Acad. Sci. USA* 80:726-730.

Alternatively, optimal alignment of sequences for comparison may be conducted by the local identity algorithm of Smith and Waterman (1981) *Add. APL. Math* 2:482, by the identity alignment algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48:443, by the search for similarity methods of Pearson and Lipman (1988) *Proc. Natl. Acad. Sci. USA* 85: 2444, by computerized implementations of these algorithms (GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group (GCG), 575 Science Dr., Madison, WI), or by inspection.

One preferred example of algorithms that are suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul et al. (1977) *Nucl. Acids Res.* 25:3389-3402 and Altschul et al. (1990) *J. Mol. Biol.* 215:403-410, respectively. BLAST and BLAST 2.0 can be used, for example with the parameters described herein, to determine percent sequence identity for the polynucleotides and polypeptides of the invention. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information. For amino acid sequences, a scoring matrix can be used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T and X determine the sensitivity and speed of the alignment.

In one preferred approach, the "percentage of sequence identity" is determined by comparing two optimally aligned sequences over a window of comparison of at least 20 positions, wherein the portion of the polypeptide sequence in the comparison window may comprise additions or deletions (*i.e.*, gaps) of 20 percent or less, usually 5 to 15 percent, or 10 to 12 percent, as compared to the reference sequences (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the reference sequence (*i.e.*, the window size) and multiplying the results by 100 to yield the percentage of sequence identity.

Within other illustrative embodiments, a polypeptide may be a xenogeneic polypeptide that comprises an polypeptide having substantial sequence identity, as described above, to the human polypeptide (also termed autologous antigen) which served as a reference polypeptide, but which xenogeneic polypeptide is derived from a different, non-human species. One skilled in the art will recognize that "self" antigens are often poor stimulators of CD8+ and CD4+ T-lymphocyte responses,

and therefore efficient immunotherapeutic strategies directed against tumor polypeptides require the development of methods to overcome immune tolerance to particular self tumor polypeptides. For example, humans immunized with prostate protein from a xenogeneic (non human) origin are capable of mounting an immune response against the counterpart human protein, *e.g.* the human prostate tumor protein present on human tumor cells. Accordingly, the present invention provides methods for purifying the xenogeneic form of the tumor proteins set forth herein, such as the polypeptides set forth in SEQ ID NOs:235, 237, and 245, or those encoded by polynucleotide sequences set forth in SEQ ID NOs:1-234, 236, and 244.

10 Therefore, one aspect of the present invention provides xenogeneic variants of the polypeptide compositions described herein. Such xenogeneic variants generally encompassed by the present invention will typically exhibit at least about 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% or more identity along their lengths, to a polypeptide sequences set forth herein.

15 More particularly, the invention is directed to mouse, rat, monkey, porcine and other non-human polypeptides which can be used as xenogeneic forms of human polypeptides set forth herein, to induce immune responses directed against tumor polypeptides of the invention.

Within other illustrative embodiments, a polypeptide may be a fusion polypeptide that comprises multiple polypeptides as described herein, or that comprises at least one polypeptide as described herein and an unrelated sequence, such as a known tumor protein. A fusion partner may, for example, assist in providing T helper epitopes (an immunological fusion partner), preferably T helper epitopes recognized by humans, or may assist in expressing the protein (an expression enhancer) at higher yields than the native recombinant protein. Certain preferred fusion partners are both immunological and expression enhancing fusion partners. Other fusion partners may be selected so as to increase the solubility of the polypeptide or to enable the polypeptide to be targeted to desired intracellular compartments. Still further fusion partners include affinity tags, which facilitate purification of the polypeptide.

30 Fusion polypeptides may generally be prepared using standard techniques, including chemical conjugation. Preferably, a fusion polypeptide is

expressed as a recombinant polypeptide, allowing the production of increased levels, relative to a non-fused polypeptide, in an expression system. Briefly, DNA sequences encoding the polypeptide components may be assembled separately, and ligated into an appropriate expression vector. The 3' end of the DNA sequence encoding one polypeptide component is ligated, with or without a peptide linker, to the 5' end of a DNA sequence encoding the second polypeptide component so that the reading frames of the sequences are in phase. This permits translation into a single fusion polypeptide that retains the biological activity of both component polypeptides.

A peptide linker sequence may be employed to separate the first and second polypeptide components by a distance sufficient to ensure that each polypeptide folds into its secondary and tertiary structures. Such a peptide linker sequence is incorporated into the fusion polypeptide using standard techniques well known in the art. Suitable peptide linker sequences may be chosen based on the following factors: (1) their ability to adopt a flexible extended conformation; (2) their inability to adopt a secondary structure that could interact with functional epitopes on the first and second polypeptides; and (3) the lack of hydrophobic or charged residues that might react with the polypeptide functional epitopes. Preferred peptide linker sequences contain Gly, Asn and Ser residues. Other near neutral amino acids, such as Thr and Ala may also be used in the linker sequence. Amino acid sequences which may be usefully employed as linkers include those disclosed in Maratea et al., *Gene* 40:39-46, 1985; Murphy et al., *Proc. Natl. Acad. Sci. USA* 83:8258-8262, 1986; U.S. Patent No. 4,935,233 and U.S. Patent No. 4,751,180. The linker sequence may generally be from 1 to about 50 amino acids in length. Linker sequences are not required when the first and second polypeptides have non-essential N-terminal amino acid regions that can be used to separate the functional domains and prevent steric interference.

The ligated DNA sequences are operably linked to suitable transcriptional or translational regulatory elements. The regulatory elements responsible for expression of DNA are located only 5' to the DNA sequence encoding the first polypeptides. Similarly, stop codons required to end translation and transcription termination signals are only present 3' to the DNA sequence encoding the second polypeptide.

The fusion polypeptide can comprise a polypeptide as described herein together with an unrelated immunogenic protein, such as an immunogenic protein capable of eliciting a recall response. Examples of such proteins include tetanus, tuberculosis and hepatitis proteins (*see, for example, Stoute et al. New Engl. J. Med.*, 5 336:86-91, 1997).

In one preferred embodiment, the immunological fusion partner is derived from a *Mycobacterium* sp., such as a *Mycobacterium tuberculosis*-derived Ra12 fragment. Ra12 compositions and methods for their use in enhancing the expression and/or immunogenicity of heterologous polynucleotide/polypeptide sequences is 10 described in U.S. Patent Application 60/158,585, the disclosure of which is incorporated herein by reference in its entirety. Briefly, Ra12 refers to a polynucleotide region that is a subsequence of a *Mycobacterium tuberculosis* MTB32A nucleic acid. MTB32A is a serine protease of 32 KD molecular weight encoded by a gene in virulent and avirulent strains of *M. tuberculosis*. The nucleotide sequence and amino acid 15 sequence of MTB32A have been described (for example, U.S. Patent Application 60/158,585; *see also, Skeiky et al., Infection and Immun.* (1999) 67:3998-4007, incorporated herein by reference). C-terminal fragments of the MTB32A coding sequence express at high levels and remain as a soluble polypeptides throughout the purification process. Moreover, Ra12 may enhance the immunogenicity of heterologous 20 immunogenic polypeptides with which it is fused. One preferred Ra12 fusion polypeptide comprises a 14 KD C-terminal fragment corresponding to amino acid residues 192 to 323 of MTB32A. Other preferred Ra12 polynucleotides generally comprise at least about 15 consecutive nucleotides, at least about 30 nucleotides, at least about 60 nucleotides, at least about 100 nucleotides, at least about 200 nucleotides, 25 or at least about 300 nucleotides that encode a portion of a Ra12 polypeptide. Ra12 polynucleotides may comprise a native sequence (*i.e.*, an endogenous sequence that encodes a Ra12 polypeptide or a portion thereof) or may comprise a variant of such a sequence. Ra12 polynucleotide variants may contain one or more substitutions, additions, deletions and/or insertions such that the biological activity of the encoded 30 fusion polypeptide is not substantially diminished, relative to a fusion polypeptide comprising a native Ra12 polypeptide. Variants preferably exhibit at least about 70%

identity, more preferably at least about 80% identity and most preferably at least about 90% identity to a polynucleotide sequence that encodes a native Ra12 polypeptide or a portion thereof.

Within other preferred embodiments, an immunological fusion partner is
5 derived from protein D, a surface protein of the gram-negative bacterium *Haemophilus influenza B* (WO 91/18926). Preferably, a protein D derivative comprises approximately the first third of the protein (e.g., the first N-terminal 100-110 amino acids), and a protein D derivative may be lipidated. Within certain preferred
10 embodiments, the first 109 residues of a Lipoprotein D fusion partner is included on the N-terminus to provide the polypeptide with additional exogenous T-cell epitopes and to increase the expression level in *E. coli* (thus functioning as an expression enhancer). The lipid tail ensures optimal presentation of the antigen to antigen presenting cells. Other fusion partners include the non-structural protein from influenzae virus, NS1 (hemagglutinin). Typically, the N-terminal 81 amino acids are used, although different
15 fragments that include T-helper epitopes may be used.

In another embodiment, the immunological fusion partner is the protein known as LYTA, or a portion thereof (preferably a C-terminal portion). LYTA is derived from *Streptococcus pneumoniae*, which synthesizes an N-acetyl-L-alanine amidase known as amidase LYTA (encoded by the *LytA* gene; *Gene* 43:265-292,
20 1986). LYTA is an autolysin that specifically degrades certain bonds in the peptidoglycan backbone. The C-terminal domain of the LYTA protein is responsible for the affinity to the choline or to some choline analogues such as DEAE. This property has been exploited for the development of *E. coli* C-LYTA expressing plasmids useful for expression of fusion proteins. Purification of hybrid proteins
25 containing the C-LYTA fragment at the amino terminus has been described (see *Biotechnology* 10:795-798, 1992). Within a preferred embodiment, a repeat portion of LYTA may be incorporated into a fusion polypeptide. A repeat portion is found in the C-terminal region starting at residue 178. A particularly preferred repeat portion incorporates residues 188-305.

30 Yet another illustrative embodiment involves fusion polypeptides, and the polynucleotides encoding them, wherein the fusion partner comprises a targeting

signal capable of directing a polypeptide to the endosomal/lysosomal compartment, as described in U.S. Patent No. 5,633,234. An immunogenic polypeptide of the invention, when fused with this targeting signal, will associate more efficiently with MHC class II molecules and thereby provide enhanced in vivo stimulation of CD4⁺ T-cells specific for the polypeptide.

Polypeptides of the invention are prepared using any of a variety of well known synthetic and/or recombinant techniques, the latter of which are further described below. Polypeptides, portions and other variants generally less than about 150 amino acids can be generated by synthetic means, using techniques well known to those of ordinary skill in the art. In one illustrative example, such polypeptides are synthesized using any of the commercially available solid-phase techniques, such as the Merrifield solid-phase synthesis method, where amino acids are sequentially added to a growing amino acid chain. See Merrifield, *J. Am. Chem. Soc.* 85:2149-2146, 1963. Equipment for automated synthesis of polypeptides is commercially available from suppliers such as Perkin Elmer/Applied BioSystems Division (Foster City, CA), and may be operated according to the manufacturer's instructions.

In general, polypeptide compositions (including fusion polypeptides) of the invention are isolated. An "isolated" polypeptide is one that is removed from its original environment. For example, a naturally-occurring protein or polypeptide is isolated if it is separated from some or all of the coexisting materials in the natural system. Preferably, such polypeptides are also purified, e.g., are at least about 90% pure, more preferably at least about 95% pure and most preferably at least about 99% pure.

POLYNUCLEOTIDE COMPOSITIONS

The present invention, in other aspects, provides polynucleotide compositions. The terms "DNA" and "polynucleotide" are used essentially interchangeably herein to refer to a DNA molecule that has been isolated free of total genomic DNA of a particular species. "Isolated," as used herein, means that a polynucleotide is substantially away from other coding sequences, and that the DNA molecule does not contain large portions of unrelated coding DNA, such as large

chromosomal fragments or other functional genes or polypeptide coding regions. Of course, this refers to the DNA molecule as originally isolated, and does not exclude genes or coding regions later added to the segment by the hand of man.

As will be understood by those skilled in the art, the polynucleotide
5 compositions of this invention can include genomic sequences, extra-genomic and plasmid-encoded sequences and smaller engineered gene segments that express, or may be adapted to express, proteins, polypeptides, peptides and the like. Such segments may be naturally isolated, or modified synthetically by the hand of man.

As will be also recognized by the skilled artisan, polynucleotides of the
10 invention may be single-stranded (coding or antisense) or double-stranded, and may be DNA (genomic, cDNA or synthetic) or RNA molecules. RNA molecules may include HnRNA molecules, which contain introns and correspond to a DNA molecule in a one-to-one manner, and mRNA molecules, which do not contain introns. Additional coding or non-coding sequences may, but need not, be present within a polynucleotide of the
15 present invention, and a polynucleotide may, but need not, be linked to other molecules and/or support materials.

Polynucleotides may comprise a native sequence (*i.e.*, an endogenous sequence that encodes a polypeptide/protein of the invention or a portion thereof) or may comprise a sequence that encodes a variant or derivative, preferably and
20 immunogenic variant or derivative, of such a sequence.

Therefore, according to another aspect of the present invention, polynucleotide compositions are provided that comprise some or all of a polynucleotide sequence set forth in any one of SEQ ID NOs:1-234, 236, and 244, complements of a polynucleotide sequence set forth in any one of SEQ ID NOs:1-234, 236, and 244, and
25 degenerate variants of a polynucleotide sequence set forth in any one of SEQ ID NOs:1-234, 236, and 244. In certain preferred embodiments, the polynucleotide sequences set forth herein encode immunogenic polypeptides, as described above.

In other related embodiments, the present invention provides polynucleotide variants having substantial identity to the sequences disclosed herein in
30 SEQ ID NOs:1-234, 236, and 244, for example those comprising at least 70% sequence identity, preferably at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% or

higher, sequence identity compared to a polynucleotide sequence of this invention using the methods described herein, (e.g., BLAST analysis using standard parameters, as described below). One skilled in this art will recognize that these values can be appropriately adjusted to determine corresponding identity of proteins encoded by two
5 nucleotide sequences by taking into account codon degeneracy, amino acid similarity, reading frame positioning and the like.

Typically, polynucleotide variants will contain one or more substitutions, additions, deletions and/or insertions, preferably such that the immunogenicity of the polypeptide encoded by the variant polynucleotide is not substantially diminished
10 relative to a polypeptide encoded by a polynucleotide sequence specifically set forth herein). The term "variants" should also be understood to encompass homologous genes of xenogeneic origin.

In additional embodiments, the present invention provides polynucleotide fragments comprising or consisting of various lengths of contiguous
15 stretches of sequence identical to or complementary to one or more of the sequences disclosed herein. For example, polynucleotides are provided by this invention that comprise or consist of at least about 10, 15, 20, 30, 40, 50, 75, 100, 150, 200, 300, 400, 500 or 1000 or more contiguous nucleotides of one or more of the sequences disclosed herein as well as all intermediate lengths there between. It will be readily understood
20 that "intermediate lengths", in this context, means any length between the quoted values, such as 16, 17, 18, 19, *etc.*; 21, 22, 23, *etc.*; 30, 31, 32, *etc.*; 50, 51, 52, 53, *etc.*; 100, 101, 102, 103, *etc.*; 150, 151, 152, 153, *etc.*; including all integers through 200-500; 500-1,000, and the like. A polynucleotide sequence as described here may be extended at one or both ends by additional nucleotides not found in the native sequence.
25 This additional sequence may consist of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides at either end of the disclosed sequence or at both ends of the disclosed sequence.

In another embodiment of the invention, polynucleotide compositions are provided that are capable of hybridizing under moderate to high stringency
30 conditions to a polynucleotide sequence provided herein, or a fragment thereof, or a complementary sequence thereof. Hybridization techniques are well known in the art

of molecular biology. For purposes of illustration, suitable moderately stringent conditions for testing the hybridization of a polynucleotide of this invention with other polynucleotides include prewashing in a solution of 5 X SSC, 0.5% SDS, 1.0 mM EDTA (pH 8.0); hybridizing at 50°C-60°C, 5 X SSC, overnight; followed by washing
5 twice at 65°C for 20 minutes with each of 2X, 0.5X and 0.2X SSC containing 0.1% SDS. One skilled in the art will understand that the stringency of hybridization can be readily manipulated, such as by altering the salt content of the hybridization solution and/or the temperature at which the hybridization is performed. For example, in another embodiment, suitable highly stringent hybridization conditions include those
10 described above, with the exception that the temperature of hybridization is increased, e.g., to 60-65°C or 65-70°C.

In certain preferred embodiments, the polynucleotides described above, e.g., polynucleotide variants, fragments and hybridizing sequences, encode polypeptides that are immunologically cross-reactive with a polypeptide sequence
15 specifically set forth herein. In other preferred embodiments, such polynucleotides encode polypeptides that have a level of immunogenic activity of at least about 50%, preferably at least about 70%, and more preferably at least about 90% of that for a polypeptide sequence specifically set forth herein.

The polynucleotides of the present invention, or fragments thereof,
20 regardless of the length of the coding sequence itself, may be combined with other DNA sequences, such as promoters, polyadenylation signals, additional restriction enzyme sites, multiple cloning sites, other coding segments, and the like, such that their overall length may vary considerably. It is therefore contemplated that a nucleic acid fragment of almost any length may be employed, with the total length preferably being
25 limited by the ease of preparation and use in the intended recombinant DNA protocol. For example, illustrative polynucleotide segments with total lengths of about 10,000, about 5000, about 3000, about 2,000, about 1,000, about 500, about 200, about 100, about 50 base pairs in length, and the like, (including all intermediate lengths) are contemplated to be useful in many implementations of this invention.

30 When comparing polynucleotide sequences, two sequences are said to be "identical" if the sequence of nucleotides in the two sequences is the same when aligned

for maximum correspondence, as described below. Comparisons between two sequences are typically performed by comparing the sequences over a comparison window to identify and compare local regions of sequence similarity. A "comparison window" as used herein, refers to a segment of at least about 20 contiguous positions, usually 30 to about 75, 40 to about 50, in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned.

Optimal alignment of sequences for comparison may be conducted using the Megalign program in the Lasergene suite of bioinformatics software (DNASTAR, Inc., Madison, WI), using default parameters. This program embodies several alignment schemes described in the following references: Dayhoff, M.O. (1978) A model of evolutionary change in proteins – Matrices for detecting distant relationships. In Dayhoff, M.O. (ed.) *Atlas of Protein Sequence and Structure*, National Biomedical Research Foundation, Washington DC Vol. 5, Suppl. 3, pp. 345-358; Hein J. (1990) Unified Approach to Alignment and Phylogenies pp. 626-645 *Methods in Enzymology* vol. 183, Academic Press, Inc., San Diego, CA; Higgins, D.G. and Sharp, P.M. (1989) *CABIOS* 5:151-153; Myers, E.W. and Muller W. (1988) *CABIOS* 4:11-17; Robinson, E.D. (1971) *Comb. Theor* 11:105; Santou, N. Nes, M. (1987) *Mol. Biol. Evol.* 4:406-425; Sneath, P.H.A. and Sokal, R.R. (1973) *Numerical Taxonomy – the Principles and Practice of Numerical Taxonomy*, Freeman Press, San Francisco, CA; Wilbur, W.J. and Lipman, D.J. (1983) *Proc. Natl. Acad., Sci. USA* 80:726-730.

Alternatively, optimal alignment of sequences for comparison may be conducted by the local identity algorithm of Smith and Waterman (1981) *Add. APL. Math* 2:482, by the identity alignment algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48:443, by the search for similarity methods of Pearson and Lipman (1988) *Proc. Natl. Acad. Sci. USA* 85: 2444, by computerized implementations of these algorithms (GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group (GCG), 575 Science Dr., Madison, WI), or by inspection.

One preferred example of algorithms that are suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0

algorithms, which are described in Altschul et al. (1977) *Nucl. Acids Res.* 25:3389-3402 and Altschul et al. (1990) *J. Mol. Biol.* 215:403-410, respectively. BLAST and BLAST 2.0 can be used, for example with the parameters described herein, to determine percent sequence identity for the polynucleotides of the invention. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information. In one illustrative example, cumulative scores can be calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always >0) and N (penalty score for mismatching residues; always <0). Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, and expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff and Henikoff (1989) *Proc. Natl. Acad. Sci. USA* 89:10915) alignments, (B) of 50, expectation (E) of 10, M=5, N=-4 and a comparison of both strands.

Preferably, the "percentage of sequence identity" is determined by comparing two optimally aligned sequences over a window of comparison of at least 20 positions, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (*i.e.*, gaps) of 20 percent or less, usually 5 to 15 percent, or 10 to 12 percent, as compared to the reference sequences (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid bases occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the reference sequence (*i.e.*, the window size) and multiplying the results by 100 to yield the percentage of sequence identity.

It will be appreciated by those of ordinary skill in the art that, as a result of the degeneracy of the genetic code, there are many nucleotide sequences that encode a polypeptide as described herein. Some of these polynucleotides bear minimal

homology to the nucleotide sequence of any native gene. Nonetheless, polynucleotides that vary due to differences in codon usage are specifically contemplated by the present invention. Further, alleles of the genes comprising the polynucleotide sequences provided herein are within the scope of the present invention. Alleles are endogenous
5 genes that are altered as a result of one or more mutations, such as deletions, additions and/or substitutions of nucleotides. The resulting mRNA and protein may, but need not, have an altered structure or function. Alleles may be identified using standard techniques (such as hybridization, amplification and/or database sequence comparison).

Therefore, in another embodiment of the invention, a mutagenesis
10 approach, such as site-specific mutagenesis, is employed for the preparation of immunogenic variants and/or derivatives of the polypeptides described herein. By this approach, specific modifications in a polypeptide sequence can be made through mutagenesis of the underlying polynucleotides that encode them. These techniques provides a straightforward approach to prepare and test sequence variants, for example,
15 incorporating one or more of the foregoing considerations, by introducing one or more nucleotide sequence changes into the polynucleotide.

Site-specific mutagenesis allows the production of mutants through the use of specific oligonucleotide sequences which encode the DNA sequence of the desired mutation, as well as a sufficient number of adjacent nucleotides, to provide a
20 primer sequence of sufficient size and sequence complexity to form a stable duplex on both sides of the deletion junction being traversed. Mutations may be employed in a selected polynucleotide sequence to improve, alter, decrease, modify, or otherwise change the properties of the polynucleotide itself, and/or alter the properties, activity, composition, stability, or primary sequence of the encoded polypeptide.

25 In certain embodiments of the present invention, the inventors contemplate the mutagenesis of the disclosed polynucleotide sequences to alter one or more properties of the encoded polypeptide, such as the immunogenicity of a polypeptide vaccine. The techniques of site-specific mutagenesis are well-known in the art, and are widely used to create variants of both polypeptides and polynucleotides.
30 For example, site-specific mutagenesis is often used to alter a specific portion of a DNA molecule. In such embodiments, a primer comprising typically about 14 to about 25

nucleotides or so in length is employed, with about 5 to about 10 residues on both sides of the junction of the sequence being altered.

As will be appreciated by those of skill in the art, site-specific mutagenesis techniques have often employed a phage vector that exists in both a single stranded and double stranded form. Typical vectors useful in site-directed mutagenesis include vectors such as the M13 phage. These phage are readily commercially-available and their use is generally well-known to those skilled in the art. Double-stranded plasmids are also routinely employed in site directed mutagenesis that eliminates the step of transferring the gene of interest from a plasmid to a phage.

10 In general, site-directed mutagenesis in accordance herewith is performed by first obtaining a single-stranded vector or melting apart of two strands of a double-stranded vector that includes within its sequence a DNA sequence that encodes the desired peptide. An oligonucleotide primer bearing the desired mutated sequence is prepared, generally synthetically. This primer is then annealed with the
15 single-stranded vector, and subjected to DNA polymerizing enzymes such as *E. coli* polymerase I Klenow fragment, in order to complete the synthesis of the mutation-bearing strand. Thus, a heteroduplex is formed wherein one strand encodes the original non-mutated sequence and the second strand bears the desired mutation. This heteroduplex vector is then used to transform appropriate cells, such as *E. coli* cells, and
20 clones are selected which include recombinant vectors bearing the mutated sequence arrangement.

The preparation of sequence variants of the selected peptide-encoding DNA segments using site-directed mutagenesis provides a means of producing potentially useful species and is not meant to be limiting as there are other ways in
25 which sequence variants of peptides and the DNA sequences encoding them may be obtained. For example, recombinant vectors encoding the desired peptide sequence may be treated with mutagenic agents, such as hydroxylamine, to obtain sequence variants. Specific details regarding these methods and protocols are found in the teachings of Maloy *et al.*, 1994; Segal, 1976; Prokop and Bajpai, 1991; Kuby, 1994;
30 and Maniatis *et al.*, 1982, each incorporated herein by reference, for that purpose.

As used herein, the term "oligonucleotide directed mutagenesis procedure" refers to template-dependent processes and vector-mediated propagation which result in an increase in the concentration of a specific nucleic acid molecule relative to its initial concentration, or in an increase in the concentration of a detectable
5 signal, such as amplification. As used herein, the term "oligonucleotide directed mutagenesis procedure" is intended to refer to a process that involves the template-dependent extension of a primer molecule. The term template dependent process refers to nucleic acid synthesis of an RNA or a DNA molecule wherein the sequence of the newly synthesized strand of nucleic acid is dictated by the well-known
10 rules of complementary base pairing (see, for example, Watson, 1987). Typically, vector mediated methodologies involve the introduction of the nucleic acid fragment into a DNA or RNA vector, the clonal amplification of the vector, and the recovery of the amplified nucleic acid fragment. Examples of such methodologies are provided by U. S. Patent No. 4,237,224, specifically incorporated herein by reference in its entirety.

15 In another approach for the production of polypeptide variants of the present invention, recursive sequence recombination, as described in U.S. Patent No. 5,837,458, may be employed. In this approach, iterative cycles of recombination and screening or selection are performed to "evolve" individual polynucleotide variants of the invention having, for example, enhanced immunogenic activity.

20 In other embodiments of the present invention, the polynucleotide sequences provided herein can be advantageously used as probes or primers for nucleic acid hybridization. As such, it is contemplated that nucleic acid segments that comprise or consist of a sequence region of at least about a 15 nucleotide long contiguous sequence that has the same sequence as, or is complementary to, a 15 nucleotide long
25 contiguous sequence disclosed herein will find particular utility. Longer contiguous identical or complementary sequences, e.g., those of about 20, 30, 40, 50, 100, 200, 500, 1000 (including all intermediate lengths) and even up to full length sequences will also be of use in certain embodiments.

The ability of such nucleic acid probes to specifically hybridize to a
30 sequence of interest will enable them to be of use in detecting the presence of complementary sequences in a given sample. However, other uses are also envisioned,

such as the use of the sequence information for the preparation of mutant species primers, or primers for use in preparing other genetic constructions.

Polynucleotide molecules having sequence regions consisting of contiguous nucleotide stretches of 10-14, 15-20, 30, 50, or even of 100-200 nucleotides or so (including intermediate lengths as well), identical or complementary to a polynucleotide sequence disclosed herein, are particularly contemplated as hybridization probes for use in, *e.g.*, Southern and Northern blotting. This would allow a gene product, or fragment thereof, to be analyzed, both in diverse cell types and also in various bacterial cells. The total size of fragment, as well as the size of the complementary stretch(es), will ultimately depend on the intended use or application of the particular nucleic acid segment. Smaller fragments will generally find use in hybridization embodiments, wherein the length of the contiguous complementary region may be varied, such as between about 15 and about 100 nucleotides, but larger contiguous complementarity stretches may be used, according to the length complementary sequences one wishes to detect.

The use of a hybridization probe of about 15-25 nucleotides in length allows the formation of a duplex molecule that is both stable and selective. Molecules having contiguous complementary sequences over stretches greater than 15 bases in length are generally preferred, though, in order to increase stability and selectivity of the hybrid, and thereby improve the quality and degree of specific hybrid molecules obtained. One will generally prefer to design nucleic acid molecules having gene-complementary stretches of 15 to 25 contiguous nucleotides, or even longer where desired.

Hybridization probes may be selected from any portion of any of the sequences disclosed herein. All that is required is to review the sequences set forth herein, or to any continuous portion of the sequences, from about 15-25 nucleotides in length up to and including the full length sequence, that one wishes to utilize as a probe or primer. The choice of probe and primer sequences may be governed by various factors. For example, one may wish to employ primers from towards the termini of the total sequence.

Small polynucleotide segments or fragments may be readily prepared by, for example, directly synthesizing the fragment by chemical means, as is commonly practiced using an automated oligonucleotide synthesizer. Also, fragments may be obtained by application of nucleic acid reproduction technology, such as the PCR™
5 technology of U. S. Patent 4,683,202 (incorporated herein by reference), by introducing selected sequences into recombinant vectors for recombinant production, and by other recombinant DNA techniques generally known to those of skill in the art of molecular biology.

The nucleotide sequences of the invention may be used for their ability
10 to selectively form duplex molecules with complementary stretches of the entire gene or gene fragments of interest. Depending on the application envisioned, one will typically desire to employ varying conditions of hybridization to achieve varying degrees of selectivity of probe towards target sequence. For applications requiring high selectivity, one will typically desire to employ relatively stringent conditions to form
15 the hybrids, *e.g.*, one will select relatively low salt and/or high temperature conditions, such as provided by a salt concentration of from about 0.02 M to about 0.15 M salt at temperatures of from about 50°C to about 70°C. Such selective conditions tolerate little, if any, mismatch between the probe and the template or target strand, and would be particularly suitable for isolating related sequences.

20 Of course, for some applications, for example, where one desires to prepare mutants employing a mutant primer strand hybridized to an underlying template, less stringent (reduced stringency) hybridization conditions will typically be needed in order to allow formation of the heteroduplex. In these circumstances, one may desire to employ salt conditions such as those of from about 0.15 M to about 0.9 M
25 salt, at temperatures ranging from about 20°C to about 55°C. Cross-hybridizing species can thereby be readily identified as positively hybridizing signals with respect to control hybridizations. In any case, it is generally appreciated that conditions can be rendered more stringent by the addition of increasing amounts of formamide, which serves to destabilize the hybrid duplex in the same manner as increased temperature. Thus,
30 hybridization conditions can be readily manipulated, and thus will generally be a method of choice depending on the desired results.

According to another embodiment of the present invention, polynucleotide compositions comprising antisense oligonucleotides are provided. Antisense oligonucleotides have been demonstrated to be effective and targeted inhibitors of protein synthesis, and, consequently, provide a therapeutic approach by which a disease can be treated by inhibiting the synthesis of proteins that contribute to the disease. The efficacy of antisense oligonucleotides for inhibiting protein synthesis is well established. For example, the synthesis of polygalacturonase and the muscarine type 2 acetylcholine receptor are inhibited by antisense oligonucleotides directed to their respective mRNA sequences (U. S. Patent 5,739,119 and U. S. Patent 5,759,829). Further, examples of antisense inhibition have been demonstrated with the nuclear protein cyclin, the multiple drug resistance gene (MDG1), ICAM-1, E-selectin, STK-1, striatal GABA_A receptor and human EGF (Jaskulski *et al.*, Science. 1988 Jun 10;240(4858):1544-6; Vasanthakumar and Ahmed, Cancer Commun. 1989;1(4):225-32; Peris *et al.*, Brain Res Mol Brain Res. 1998 Jun 15;57(2):310-20; U. S. Patent 5,801,154; U.S. Patent 5,789,573; U. S. Patent 5,718,709 and U.S. Patent 5,610,288). Antisense constructs have also been described that inhibit and can be used to treat a variety of abnormal cellular proliferations, *e.g.* cancer (U. S. Patent 5,747,470; U. S. Patent 5,591,317 and U. S. Patent 5,783,683).

Therefore, in certain embodiments, the present invention provides oligonucleotide sequences that comprise all, or a portion of, any sequence that is capable of specifically binding to polynucleotide sequence described herein, or a complement thereof. In one embodiment, the antisense oligonucleotides comprise DNA or derivatives thereof. In another embodiment, the oligonucleotides comprise RNA or derivatives thereof. In a third embodiment, the oligonucleotides are modified DNAs comprising a phosphorothioated modified backbone. In a fourth embodiment, the oligonucleotide sequences comprise peptide nucleic acids or derivatives thereof. In each case, preferred compositions comprise a sequence region that is complementary, and more preferably substantially-complementary, and even more preferably, completely complementary to one or more portions of polynucleotides disclosed herein. Selection of antisense compositions specific for a given gene sequence is based upon analysis of the chosen target sequence and determination of secondary structure, T_m ,

binding energy, and relative stability. Antisense compositions may be selected based upon their relative inability to form dimers, hairpins, or other secondary structures that would reduce or prohibit specific binding to the target mRNA in a host cell. Highly preferred target regions of the mRNA, are those which are at or near the AUG translation initiation codon, and those sequences which are substantially complementary to 5' regions of the mRNA. These secondary structure analyses and target site selection considerations can be performed, for example, using v.4 of the OLIGO primer analysis software and/or the BLASTN 2.0.5 algorithm software (Altschul *et al.*, Nucleic Acids Res. 1997, 25(17):3389-402).

10 The use of an antisense delivery method employing a short peptide vector, termed MPG (27 residues), is also contemplated. The MPG peptide contains a hydrophobic domain derived from the fusion sequence of HIV gp41 and a hydrophilic domain from the nuclear localization sequence of SV40 T-antigen (Morris *et al.*, Nucleic Acids Res. 1997 Jul 15;25(14):2730-6). It has been demonstrated that several
15 molecules of the MPG peptide coat the antisense oligonucleotides and can be delivered into cultured mammalian cells in less than 1 hour with relatively high efficiency (90%). Further, the interaction with MPG strongly increases both the stability of the oligonucleotide to nuclease and the ability to cross the plasma membrane.

 According to another embodiment of the invention, the polynucleotide
20 compositions described herein are used in the design and preparation of ribozyme molecules for inhibiting expression of the tumor polypeptides and proteins of the present invention in tumor cells. Ribozymes are RNA-protein complexes that cleave nucleic acids in a site-specific fashion. Ribozymes have specific catalytic domains that possess endonuclease activity (Kim and Cech, Proc Natl Acad Sci U S A. 1987
25 Dec;84(24):8788-92; Forster and Symons, Cell. 1987 Apr 24;49(2):211-20). For example, a large number of ribozymes accelerate phosphoester transfer reactions with a high degree of specificity, often cleaving only one of several phosphoesters in an oligonucleotide substrate (Cech *et al.*, Cell. 1981 Dec;27(3 Pt 2):487-96; Michel and Westhof, J Mol Biol. 1990 Dec 5;216(3):585-610; Reinhold-Hurek and Shub, Nature.
30 1992 May 14;357(6374):173-6). This specificity has been attributed to the requirement

that the substrate bind via specific base-pairing interactions to the internal guide sequence ("IGS") of the ribozyme prior to chemical reaction.

Six basic varieties of naturally-occurring enzymatic RNAs are known presently. Each can catalyze the hydrolysis of RNA phosphodiester bonds *in trans* (and thus can cleave other RNA molecules) under physiological conditions. In general, enzymatic nucleic acids act by first binding to a target RNA. Such binding occurs through the target binding portion of a enzymatic nucleic acid which is held in close proximity to an enzymatic portion of the molecule that acts to cleave the target RNA. Thus, the enzymatic nucleic acid first recognizes and then binds a target RNA through complementary base-pairing, and once bound to the correct site, acts enzymatically to cut the target RNA. Strategic cleavage of such a target RNA will destroy its ability to direct synthesis of an encoded protein. After an enzymatic nucleic acid has bound and cleaved its RNA target, it is released from that RNA to search for another target and can repeatedly bind and cleave new targets.

The enzymatic nature of a ribozyme is advantageous over many technologies, such as antisense technology (where a nucleic acid molecule simply binds to a nucleic acid target to block its translation) since the concentration of ribozyme necessary to affect a therapeutic treatment is lower than that of an antisense oligonucleotide. This advantage reflects the ability of the ribozyme to act enzymatically. Thus, a single ribozyme molecule is able to cleave many molecules of target RNA. In addition, the ribozyme is a highly specific inhibitor, with the specificity of inhibition depending not only on the base pairing mechanism of binding to the target RNA, but also on the mechanism of target RNA cleavage. Single mismatches, or base-substitutions, near the site of cleavage can completely eliminate catalytic activity of a ribozyme. Similar mismatches in antisense molecules do not prevent their action (Woelf *et al.*, Proc Natl Acad Sci U S A. 1992 Aug 15;89(16):7305-9). Thus, the specificity of action of a ribozyme is greater than that of an antisense oligonucleotide binding the same RNA site.

The enzymatic nucleic acid molecule may be formed in a hammerhead, hairpin, a hepatitis δ virus, group I intron or RNaseP RNA (in association with an RNA guide sequence) or Neurospora VS RNA motif. Examples of hammerhead motifs are

described by Rossi *et al.* Nucleic Acids Res. 1992 Sep 11;20(17):4559-65. Examples of hairpin motifs are described by Hampel *et al.* (Eur. Pat. Appl. Publ. No. EP 0360257), Hampel and Tritz, Biochemistry 1989 Jun 13;28(12):4929-33; Hampel *et al.*, Nucleic Acids Res. 1990 Jan 25;18(2):299-304 and U. S. Patent 5,631,359. An example of the hepatitis δ virus motif is described by Perrotta and Been, Biochemistry. 1992 Dec 1;31(47):11843-52; an example of the RNaseP motif is described by Guerrier-Takada *et al.*, Cell. 1983 Dec;35(3 Pt 2):849-57; Neurospora VS RNA ribozyme motif is described by Collins (Saville and Collins, Cell. 1990 May 18;61(4):685-96; Saville and Collins, Proc Natl Acad Sci U S A. 1991 Oct 1;88(19):8826-30; Collins and Olive, Biochemistry. 1993 Mar 23;32(11):2795-9); and an example of the Group I intron is described in (U. S. Patent 4,987,071). All that is important in an enzymatic nucleic acid molecule of this invention is that it has a specific substrate binding site which is complementary to one or more of the target gene RNA regions, and that it have nucleotide sequences within or surrounding that substrate binding site which impart an RNA cleaving activity to the molecule. Thus the ribozyme constructs need not be limited to specific motifs mentioned herein.

Ribozymes may be designed as described in Int. Pat. Appl. Publ. No. WO 93/23569 and Int. Pat. Appl. Publ. No. WO 94/02595, each specifically incorporated herein by reference) and synthesized to be tested *in vitro* and *in vivo*, as described. Such ribozymes can also be optimized for delivery. While specific examples are provided, those in the art will recognize that equivalent RNA targets in other species can be utilized when necessary.

Ribozyme activity can be optimized by altering the length of the ribozyme binding arms, or chemically synthesizing ribozymes with modifications that prevent their degradation by serum ribonucleases (see *e.g.*, Int. Pat. Appl. Publ. No. WO 92/07065; Int. Pat. Appl. Publ. No. WO 93/15187; Int. Pat. Appl. Publ. No. WO 91/03162; Eur. Pat. Appl. Publ. No. 92110298.4; U. S. Patent 5,334,711; and Int. Pat. Appl. Publ. No. WO 94/13688, which describe various chemical modifications that can be made to the sugar moieties of enzymatic RNA molecules), modifications which enhance their efficacy in cells, and removal of stem II bases to shorten RNA synthesis times and reduce chemical requirements.

Sullivan *et al.* (Int. Pat. Appl. Publ. No. WO 94/02595) describes the general methods for delivery of enzymatic RNA molecules. Ribozymes may be administered to cells by a variety of methods known to those familiar to the art, including, but not restricted to, encapsulation in liposomes, by iontophoresis, or by
5 incorporation into other vehicles, such as hydrogels, cyclodextrins, biodegradable nanocapsules, and bioadhesive microspheres. For some indications, ribozymes may be directly delivered *ex vivo* to cells or tissues with or without the aforementioned vehicles. Alternatively, the RNA/vehicle combination may be locally delivered by direct inhalation, by direct injection or by use of a catheter, infusion pump or stent.
10 Other routes of delivery include, but are not limited to, intravascular, intramuscular, subcutaneous or joint injection, aerosol inhalation, oral (tablet or pill form), topical, systemic, ocular, intraperitoneal and/or intrathecal delivery. More detailed descriptions of ribozyme delivery and administration are provided in Int. Pat. Appl. Publ. No. WO 94/02595 and Int. Pat. Appl. Publ. No. WO 93/23569, each specifically incorporated
15 herein by reference.

Another means of accumulating high concentrations of a ribozyme(s) within cells is to incorporate the ribozyme-encoding sequences into a DNA expression vector. Transcription of the ribozyme sequences are driven from a promoter for eukaryotic RNA polymerase I (pol I), RNA polymerase II (pol II), or RNA polymerase
20 III (pol III). Transcripts from pol II or pol III promoters will be expressed at high levels in all cells; the levels of a given pol II promoter in a given cell type will depend on the nature of the gene regulatory sequences (enhancers, silencers, *etc.*) present nearby. Prokaryotic RNA polymerase promoters may also be used, providing that the prokaryotic RNA polymerase enzyme is expressed in the appropriate cells. Ribozymes
25 expressed from such promoters have been shown to function in mammalian cells. Such transcription units can be incorporated into a variety of vectors for introduction into mammalian cells, including but not restricted to, plasmid DNA vectors, viral DNA vectors (such as adenovirus or adeno-associated vectors), or viral RNA vectors (such as retroviral, semliki forest virus, sindbis virus vectors).

30 In another embodiment of the invention, peptide nucleic acids (PNAs) compositions are provided. PNA is a DNA mimic in which the nucleobases are

attached to a pseudopeptide backbone (Good and Nielsen, *Antisense Nucleic Acid Drug Dev.* 1997, 7(4) 431-37). PNA is able to be utilized in a number of methods that traditionally have used RNA or DNA. Often PNA sequences perform better in techniques than the corresponding RNA or DNA sequences and have utilities that are not inherent to RNA or DNA. A review of PNA including methods of making, characteristics of, and methods of using, is provided by Corey (*Trends Biotechnol* 1997 Jun;15(6):224-9). As such, in certain embodiments, one may prepare PNA sequences that are complementary to one or more portions of the ACE mRNA sequence, and such PNA compositions may be used to regulate, alter, decrease, or reduce the translation of ACE-specific mRNA, and thereby alter the level of ACE activity in a host cell to which such PNA compositions have been administered.

PNAs have 2-aminoethyl-glycine linkages replacing the normal phosphodiester backbone of DNA (Nielsen *et al.*, *Science* 1991 Dec 6;254(5037):1497-500; Hanvey *et al.*, *Science* 1992 Nov 27;258(5087):1481-5; Hyrup and Nielsen, *Bioorg Med Chem.* 1996 Jan;4(1):5-23). This chemistry has three important consequences: firstly, in contrast to DNA or phosphorothioate oligonucleotides, PNAs are neutral molecules; secondly, PNAs are achiral, which avoids the need to develop a stereoselective synthesis; and thirdly, PNA synthesis uses standard Boc or Fmoc protocols for solid-phase peptide synthesis, although other methods, including a modified Merrifield method, have been used.

PNA monomers or ready-made oligomers are commercially available from PerSeptive Biosystems (Framingham, MA). PNA syntheses by either Boc or Fmoc protocols are straightforward using manual or automated protocols (Norton *et al.*, *Bioorg Med Chem.* 1995 Apr;3(4):437-45). The manual protocol lends itself to the production of chemically modified PNAs or the simultaneous synthesis of families of closely related PNAs.

As with peptide synthesis, the success of a particular PNA synthesis will depend on the properties of the chosen sequence. For example, while in theory PNAs can incorporate any combination of nucleotide bases, the presence of adjacent purines can lead to deletions of one or more residues in the product. In expectation of this difficulty, it is suggested that, in producing PNAs with adjacent purines, one should

repeat the coupling of residues likely to be added inefficiently. This should be followed by the purification of PNAs by reverse-phase high-pressure liquid chromatography, providing yields and purity of product similar to those observed during the synthesis of peptides.

5 Modifications of PNAs for a given application may be accomplished by coupling amino acids during solid-phase synthesis or by attaching compounds that contain a carboxylic acid group to the exposed N-terminal amine. Alternatively, PNAs can be modified after synthesis by coupling to an introduced lysine or cysteine. The ease with which PNAs can be modified facilitates optimization for better solubility or
10 for specific functional requirements. Once synthesized, the identity of PNAs and their derivatives can be confirmed by mass spectrometry. Several studies have made and utilized modifications of PNAs (for example, Norton *et al.*, Bioorg Med Chem. 1995 Apr;3(4):437-45; Petersen *et al.*, J Pept Sci. 1995 May-Jun;1(3):175-83; Orum *et al.*, Biotechniques. 1995 Sep;19(3):472-80; Footer *et al.*, Biochemistry. 1996 Aug
15 20;35(33):10673-9; Griffith *et al.*, Nucleic Acids Res. 1995 Aug 11;23(15):3003-8; Pardridge *et al.*, Proc Natl Acad Sci U S A. 1995 Jun 6;92(12):5592-6; Boffa *et al.*, Proc Natl Acad Sci U S A. 1995 Mar 14;92(6):1901-5; Gambacorti-Passerini *et al.*, Blood. 1996 Aug 15;88(4):1411-7; Armitage *et al.*, Proc Natl Acad Sci U S A. 1997 Nov 11;94(23):12320-5; Seeger *et al.*, Biotechniques. 1997 Sep;23(3):512-7). U.S.
20 Patent No. 5,700,922 discusses PNA-DNA-PNA chimeric molecules and their uses in diagnostics, modulating protein in organisms, and treatment of conditions susceptible to therapeutics.

Methods of characterizing the antisense binding properties of PNAs are discussed in Rose (Anal Chem. 1993 Dec 15;65(24):3545-9) and Jensen *et al.*
25 (Biochemistry. 1997 Apr 22;36(16):5072-7). Rose uses capillary gel electrophoresis to determine binding of PNAs to their complementary oligonucleotide, measuring the relative binding kinetics and stoichiometry. Similar types of measurements were made by Jensen *et al.* using BIAcore™ technology.

Other applications of PNAs that have been described and will be
30 apparent to the skilled artisan include use in DNA strand invasion, antisense inhibition, mutational analysis, enhancers of transcription, nucleic acid purification, isolation of

transcriptionally active genes, blocking of transcription factor binding, genome cleavage, biosensors, *in situ* hybridization, and the like.

POLYNUCLEOTIDE IDENTIFICATION, CHARACTERIZATION AND EXPRESSION

Polynucleotides compositions of the present invention may be identified, prepared and/or manipulated using any of a variety of well established techniques (see generally, Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989, and other like references). For example, a polynucleotide may be identified, as described in more detail below, by screening a microarray of cDNAs for tumor-associated expression (*i.e.*, expression that is at least two fold greater in a tumor than in normal tissue, as determined using a representative assay provided herein). Such screens may be performed, for example, using the microarray technology of Affymetrix, Inc. (Santa Clara, CA) according to the manufacturer's instructions (and essentially as described by Schena et al., *Proc. Natl. Acad. Sci. USA* 93:10614-10619, 1996 and Heller et al., *Proc. Natl. Acad. Sci. USA* 94:2150-2155, 1997). Alternatively, polynucleotides may be amplified from cDNA prepared from cells expressing the proteins described herein, such as tumor cells.

Many template dependent processes are available to amplify a target sequences of interest present in a sample. One of the best known amplification methods is the polymerase chain reaction (PCRTM) which is described in detail in U.S. Patent Nos. 4,683,195, 4,683,202 and 4,800,159, each of which is incorporated herein by reference in its entirety. Briefly, in PCRTM, two primer sequences are prepared which are complementary to regions on opposite complementary strands of the target sequence. An excess of deoxynucleoside triphosphates is added to a reaction mixture along with a DNA polymerase (*e.g.*, *Taq* polymerase). If the target sequence is present in a sample, the primers will bind to the target and the polymerase will cause the primers to be extended along the target sequence by adding on nucleotides. By raising and lowering the temperature of the reaction mixture, the extended primers will dissociate from the target to form reaction products, excess primers will bind to the target and to the reaction product and the process is repeated. Preferably reverse transcription and PCRTM amplification procedure may be performed in order to quantify

the amount of mRNA amplified. Polymerase chain reaction methodologies are well known in the art.

Any of a number of other template dependent processes, many of which are variations of the PCRTM amplification technique, are readily known and available in the art. Illustratively, some such methods include the ligase chain reaction (referred to as LCR), described, for example, in Eur. Pat. Appl. Publ. No. 320,308 and U.S. Patent No. 4,883,750; Qbeta Replicase, described in PCT Intl. Pat. Appl. Publ. No. PCT/US87/00880; Strand Displacement Amplification (SDA) and Repair Chain Reaction (RCR). Still other amplification methods are described in Great Britain Pat. Appl. No. 2 202 328, and in PCT Intl. Pat. Appl. Publ. No. PCT/US89/01025. Other nucleic acid amplification procedures include transcription-based amplification systems (TAS) (PCT Intl. Pat. Appl. Publ. No. WO 88/10315), including nucleic acid sequence based amplification (NASBA) and 3SR. Eur. Pat. Appl. Publ. No. 329,822 describes a nucleic acid amplification process involving cyclically synthesizing single-stranded RNA ("ssRNA"), ssDNA, and double-stranded DNA (dsDNA). PCT Intl. Pat. Appl. Publ. No. WO 89/06700 describes a nucleic acid sequence amplification scheme based on the hybridization of a promoter/primer sequence to a target single-stranded DNA ("ssDNA") followed by transcription of many RNA copies of the sequence. Other amplification methods such as "RACE" (Frohman, 1990), and "one-sided PCR" (Ohara, 1989) are also well-known to those of skill in the art.

An amplified portion of a polynucleotide of the present invention may be used to isolate a full length gene from a suitable library (e.g., a tumor cDNA library) using well known techniques. Within such techniques, a library (cDNA or genomic) is screened using one or more polynucleotide probes or primers suitable for amplification. Preferably, a library is size-selected to include larger molecules. Random primed libraries may also be preferred for identifying 5' and upstream regions of genes. Genomic libraries are preferred for obtaining introns and extending 5' sequences.

For hybridization techniques, a partial sequence may be labeled (e.g., by nick-translation or end-labeling with ³²P) using well known techniques. A bacterial or bacteriophage library is then generally screened by hybridizing filters containing denatured bacterial colonies (or lawns containing phage plaques) with the labeled probe

(see Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989). Hybridizing colonies or plaques are selected and expanded, and the DNA is isolated for further analysis. cDNA clones may be analyzed to determine the amount of additional sequence by, for example, PCR
5 using a primer from the partial sequence and a primer from the vector. Restriction maps and partial sequences may be generated to identify one or more overlapping clones. The complete sequence may then be determined using standard techniques, which may involve generating a series of deletion clones. The resulting overlapping sequences can then be assembled into a single contiguous sequence. A full length cDNA
10 molecule can be generated by ligating suitable fragments, using well known techniques.

Alternatively, amplification techniques, such as those described above, can be useful for obtaining a full length coding sequence from a partial cDNA sequence. One such amplification technique is inverse PCR (see Triglia et al., *Nucl. Acids Res.* 16:8186, 1988), which uses restriction enzymes to generate a fragment in the
15 known region of the gene. The fragment is then circularized by intramolecular ligation and used as a template for PCR with divergent primers derived from the known region. Within an alternative approach, sequences adjacent to a partial sequence may be retrieved by amplification with a primer to a linker sequence and a primer specific to a known region. The amplified sequences are typically subjected to a second round of
20 amplification with the same linker primer and a second primer specific to the known region. A variation on this procedure, which employs two primers that initiate extension in opposite directions from the known sequence, is described in WO 96/38591. Another such technique is known as "rapid amplification of cDNA ends" or RACE. This technique involves the use of an internal primer and an external primer,
25 which hybridizes to a polyA region or vector sequence, to identify sequences that are 5' and 3' of a known sequence. Additional techniques include capture PCR (Lagerstrom et al., *PCR Methods Applic.* 1:111-19, 1991) and walking PCR (Parker et al., *Nucl. Acids. Res.* 19:3055-60, 1991). Other methods employing amplification may also be employed to obtain a full length cDNA sequence.

30 In certain instances, it is possible to obtain a full length cDNA sequence by analysis of sequences provided in an expressed sequence tag (EST) database, such as

that available from GenBank. Searches for overlapping ESTs may generally be performed using well known programs (e.g., NCBI BLAST searches), and such ESTs may be used to generate a contiguous full length sequence. Full length DNA sequences may also be obtained by analysis of genomic fragments.

5 In other embodiments of the invention, polynucleotide sequences or fragments thereof which encode polypeptides of the invention, or fusion proteins or functional equivalents thereof, may be used in recombinant DNA molecules to direct expression of a polypeptide in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences that encode substantially the same or a
10 functionally equivalent amino acid sequence may be produced and these sequences may be used to clone and express a given polypeptide.

As will be understood by those of skill in the art, it may be advantageous in some instances to produce polypeptide-encoding nucleotide sequences possessing non-naturally occurring codons. For example, codons preferred by a particular
15 prokaryotic or eukaryotic host can be selected to increase the rate of protein expression or to produce a recombinant RNA transcript having desirable properties, such as a half-life which is longer than that of a transcript generated from the naturally occurring sequence.

Moreover, the polynucleotide sequences of the present invention can be
20 engineered using methods generally known in the art in order to alter polypeptide encoding sequences for a variety of reasons, including but not limited to, alterations which modify the cloning, processing, and/or expression of the gene product. For example, DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide
25 sequences. In addition, site-directed mutagenesis may be used to insert new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, or introduce mutations, and so forth.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences may be ligated to a heterologous sequence to
30 encode a fusion protein. For example, to screen peptide libraries for inhibitors of polypeptide activity, it may be useful to encode a chimeric protein that can be

recognized by a commercially available antibody. A fusion protein may also be engineered to contain a cleavage site located between the polypeptide-encoding sequence and the heterologous protein sequence, so that the polypeptide may be cleaved and purified away from the heterologous moiety.

5 Sequences encoding a desired polypeptide may be synthesized, in whole or in part, using chemical methods well known in the art (see Caruthers, M. H. et al. (1980) *Nucl. Acids Res. Symp. Ser.* 215-223, Horn, T. et al. (1980) *Nucl. Acids Res. Symp. Ser.* 225-232). Alternatively, the protein itself may be produced using chemical methods to synthesize the amino acid sequence of a polypeptide, or a portion thereof.

10 For example, peptide synthesis can be performed using various solid-phase techniques (Roberge, J. Y. et al. (1995) *Science* 269:202-204) and automated synthesis may be achieved, for example, using the ABI 431A Peptide Synthesizer (Perkin Elmer, Palo Alto, CA).

A newly synthesized peptide may be substantially purified by

15 preparative high performance liquid chromatography (e.g., Creighton, T. (1983) *Proteins, Structures and Molecular Principles*, WH Freeman and Co., New York, N.Y.) or other comparable techniques available in the art. The composition of the synthetic peptides may be confirmed by amino acid analysis or sequencing (e.g., the Edman degradation procedure). Additionally, the amino acid sequence of a polypeptide, or any

20 part thereof, may be altered during direct synthesis and/or combined using chemical methods with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

In order to express a desired polypeptide, the nucleotide sequences encoding the polypeptide, or functional equivalents, may be inserted into appropriate

25 expression vector, i.e., a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence. Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding a polypeptide of interest and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA

30 techniques, synthetic techniques, and *in vivo* genetic recombination. Such techniques are described, for example, in Sambrook, J. et al. (1989) *Molecular Cloning*, A

Laboratory Manual, Cold Spring Harbor Press, Plainview, N.Y., and Ausubel, F. M. et al. (1989) Current Protocols in Molecular Biology, John Wiley & Sons, New York, N.Y.

A variety of expression vector/host systems may be utilized to contain
5 and express polynucleotide sequences. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (e.g., baculovirus); plant cell
10 systems transformed with virus expression vectors (e.g., cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems.

The "control elements" or "regulatory sequences" present in an expression vector are those non-translated regions of the vector--enhancers, promoters, 5' and 3' untranslated regions--which interact with host cellular proteins to carry out
15 transcription and translation. Such elements may vary in their strength and specificity. Depending on the vector system and host utilized, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible promoters such as the hybrid
20 lacZ promoter of the pBLUESCRIPT phagemid (Stratagene, La Jolla, Calif.) or pSPORT1 plasmid (Gibco BRL, Gaithersburg, MD) and the like may be used. In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are generally preferred. If it is necessary to generate a cell line that contains multiple copies of the sequence encoding a polypeptide, vectors based on SV40 or EBV may be advantageously used with an appropriate selectable marker.

25 In bacterial systems, any of a number of expression vectors may be selected depending upon the use intended for the expressed polypeptide. For example, when large quantities are needed, for example for the induction of antibodies, vectors which direct high level expression of fusion proteins that are readily purified may be used. Such vectors include, but are not limited to, the multifunctional *E. coli* cloning
30 and expression vectors such as pBLUESCRIPT (Stratagene), in which the sequence encoding the polypeptide of interest may be ligated into the vector in frame with

sequences for the amino-terminal Met and the subsequent 7 residues of .beta.-galactosidase so that a hybrid protein is produced; pIN vectors (Van Heeke, G. and S. M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509); and the like. pGEX Vectors (Promega, Madison, Wis.) may also be used to express foreign polypeptides as fusion
5 proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Proteins made in such systems may be designed to include heparin, thrombin, or factor XA protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety at
10 will.

In the yeast, *Saccharomyces cerevisiae*, a number of vectors containing constitutive or inducible promoters such as alpha factor, alcohol oxidase, and PGH may be used. For reviews, see Ausubel et al. (supra) and Grant et al. (1987) *Methods Enzymol.* 153:516-544.

15 In cases where plant expression vectors are used, the expression of sequences encoding polypeptides may be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of CaMV may be used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) *EMBO J.* 6:307-311. Alternatively, plant promoters such as the small subunit of
20 RUBISCO or heat shock promoters may be used (Coruzzi, G. et al. (1984) *EMBO J.* 3:1671-1680; Broglie, R. et al. (1984) *Science* 224:838-843; and Winter, J. et al. (1991) *Results Probl. Cell Differ.* 17:85-105). These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. Such techniques are described in a number of generally available reviews (see, for example, Hobbs, S. or
25 Murry, L. E. in McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York, N.Y.; pp. 191-196).

An insect system may also be used to express a polypeptide of interest. For example, in one such system, *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes in *Spodoptera frugiperda* cells or
30 in *Trichoplusia* larvae. The sequences encoding the polypeptide may be cloned into a non-essential region of the virus, such as the polyhedrin gene, and placed under control

of the polyhedrin promoter. Successful insertion of the polypeptide-encoding sequence will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein. The recombinant viruses may then be used to infect, for example, *S. frugiperda* cells or *Trichoplusia* larvae in which the polypeptide of interest may be expressed
5 (Engelhard, E. K. et al. (1994) *Proc. Natl. Acad. Sci.* 91 :3224-3227).

In mammalian host cells, a number of viral-based expression systems are generally available. For example, in cases where an adenovirus is used as an expression vector, sequences encoding a polypeptide of interest may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader
10 sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain a viable virus which is capable of expressing the polypeptide in infected host cells (Logan, J. and Shenk, T. (1984) *Proc. Natl. Acad. Sci.* 81:3655-3659). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells.

15 Specific initiation signals may also be used to achieve more efficient translation of sequences encoding a polypeptide of interest. Such signals include the ATG initiation codon and adjacent sequences. In cases where sequences encoding the polypeptide, its initiation codon, and upstream sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control
20 signals may be needed. However, in cases where only coding sequence, or a portion thereof, is inserted, exogenous translational control signals including the ATG initiation codon should be provided. Furthermore, the initiation codon should be in the correct reading frame to ensure translation of the entire insert. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic.
25 The efficiency of expression may be enhanced by the inclusion of enhancers which are appropriate for the particular cell system which is used, such as those described in the literature (Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162).

In addition, a host cell strain may be chosen for its ability to modulate the expression of the inserted sequences or to process the expressed protein in the
30 desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation.

Post-translational processing which cleaves a "prepro" form of the protein may also be used to facilitate correct insertion, folding and/or function. Different host cells such as CHO, COS, HeLa, MDCK, HEK293, and WI38, which have specific cellular machinery and characteristic mechanisms for such post-translational activities, may be
5 chosen to ensure the correct modification and processing of the foreign protein.

For long-term, high-yield production of recombinant proteins, stable expression is generally preferred. For example, cell lines which stably express a polynucleotide of interest may be transformed using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a
10 selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for 1-2 days in an enriched media before they are switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed
15 cells may be proliferated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase (Wigler, M. et al. (1977) *Cell* 11:223-32) and adenine phosphoribosyltransferase (Lowy, I. et al. (1990) *Cell* 22:817-23) genes which can be employed in tk.sup.- or
20 aprt.sup.- cells, respectively. Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, dhfr which confers resistance to methotrexate (Wigler, M. et al. (1980) *Proc. Natl. Acad. Sci.* 77:3567-70); npt, which confers resistance to the aminoglycosides, neomycin and G-418 (Colbere-Garapin, F. et al (1981) *J. Mol. Biol.* 150:1-14); and als or pat, which confer resistance to
25 chlorsulfuron and phosphinotricin acetyltransferase, respectively (Murry, *supra*). Additional selectable genes have been described, for example, trpB, which allows cells to utilize indole in place of tryptophan, or hisD, which allows cells to utilize histinol in place of histidine (Hartman, S. C. and R. C. Mulligan (1988) *Proc. Natl. Acad. Sci.* 85:8047-51). The use of visible markers has gained popularity with such markers as
30 anthocyanins, beta-glucuronidase and its substrate GUS, and luciferase and its substrate luciferin, being widely used not only to identify transformants, but also to quantify the

amount of transient or stable protein expression attributable to a specific vector system (Rhodes, C. A. et al. (1995) *Methods Mol. Biol.* 55:121-131).

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, its presence and expression may need to be confirmed. For example, if the sequence encoding a polypeptide is inserted within a marker gene sequence, recombinant cells containing sequences can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a polypeptide-encoding sequence under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

Alternatively, host cells that contain and express a desired polynucleotide sequence may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations and protein bioassay or immunoassay techniques which include, for example, membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein.

A variety of protocols for detecting and measuring the expression of polynucleotide-encoded products, using either polyclonal or monoclonal antibodies specific for the product are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on a given polypeptide may be preferred for some applications, but a competitive binding assay may also be employed. These and other assays are described, among other places, in Hampton, R. et al. (1990; Serological Methods, a Laboratory Manual, APS Press, St Paul, Minn.) and Maddox, D. E. et al. (1983; *J. Exp. Med.* 158:1211-1216).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides include oligolabeling, nick translation, end-labeling or PCR amplification using a labeled nucleotide. Alternatively, the sequences, or any portions

thereof may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits. Suitable reporter molecules or labels, which may be used include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with a polynucleotide sequence of interest may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a recombinant cell may be secreted or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides of the invention may be designed to contain signal sequences which direct secretion of the encoded polypeptide through a prokaryotic or eukaryotic cell membrane. Other recombinant constructions may be used to join sequences encoding a polypeptide of interest to nucleotide sequence encoding a polypeptide domain which will facilitate purification of soluble proteins. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp., Seattle, Wash.). The inclusion of cleavable linker sequences such as those specific for Factor XA or enterokinase (Invitrogen, San Diego, Calif.) between the purification domain and the encoded polypeptide may be used to facilitate purification. One such expression vector provides for expression of a fusion protein containing a polypeptide of interest and a nucleic acid encoding 6 histidine residues preceding a thioredoxin or an enterokinase cleavage site. The histidine residues facilitate purification on IMLAC (immobilized metal ion affinity chromatography) as described in Porath, J. et al. (1992, *Prot. Exp. Purif.* 3:263-281) while the enterokinase cleavage site provides a means for purifying the desired polypeptide from the fusion protein. A discussion of vectors which contain fusion proteins is provided in Kroll, D. J. et al. (1993; *DNA Cell Biol.* 12:441-453).

In addition to recombinant production methods, polypeptides of the invention, and fragments thereof, may be produced by direct peptide synthesis using solid-phase techniques (Merrifield J. (1963) *J. Am. Chem. Soc.* 85:2149-2154). Protein synthesis may be performed using manual techniques or by automation. Automated synthesis may be achieved, for example, using Applied Biosystems 431A Peptide Synthesizer (Perkin Elmer). Alternatively, various fragments may be chemically synthesized separately and combined using chemical methods to produce the full length molecule.

ANTIBODY COMPOSITIONS, FRAGMENTS THEREOF AND OTHER BINDING AGENTS

According to another aspect, the present invention further provides binding agents, such as antibodies and antigen-binding fragments thereof, that exhibit immunological binding to a tumor polypeptide disclosed herein, or to a portion, variant or derivative thereof. An antibody, or antigen-binding fragment thereof, is said to "specifically bind," "immunologically bind," and/or is "immunologically reactive" to a polypeptide of the invention if it reacts at a detectable level (within, for example, an ELISA assay) with the polypeptide, and does not react detectably with unrelated polypeptides under similar conditions.

Immunological binding, as used in this context, generally refers to the non-covalent interactions of the type which occur between an immunoglobulin molecule and an antigen for which the immunoglobulin is specific. The strength, or affinity of immunological binding interactions can be expressed in terms of the dissociation constant (K_d) of the interaction, wherein a smaller K_d represents a greater affinity. Immunological binding properties of selected polypeptides can be quantified using methods well known in the art. One such method entails measuring the rates of antigen-binding site/antigen complex formation and dissociation, wherein those rates depend on the concentrations of the complex partners, the affinity of the interaction, and on geometric parameters that equally influence the rate in both directions. Thus, both the "on rate constant" (K_{on}) and the "off rate constant" (K_{off}) can be determined by calculation of the concentrations and the actual rates of association and dissociation. The ratio of K_{off}/K_{on} enables cancellation of all parameters not related to affinity, and is

thus equal to the dissociation constant K_d . See, generally, Davies et al. (1990) Annual Rev. Biochem. 59:439-473.

An "antigen-binding site," or "binding portion" of an antibody refers to the part of the immunoglobulin molecule that participates in antigen binding. The antigen binding site is formed by amino acid residues of the N-terminal variable ("V") regions of the heavy ("H") and light ("L") chains. Three highly divergent stretches within the V regions of the heavy and light chains are referred to as "hypervariable regions" which are interposed between more conserved flanking stretches known as "framework regions," or "FRs". Thus the term "FR" refers to amino acid sequences which are naturally found between and adjacent to hypervariable regions in immunoglobulins. In an antibody molecule, the three hypervariable regions of a light chain and the three hypervariable regions of a heavy chain are disposed relative to each other in three dimensional space to form an antigen-binding surface. The antigen-binding surface is complementary to the three-dimensional surface of a bound antigen, and the three hypervariable regions of each of the heavy and light chains are referred to as "complementarity-determining regions," or "CDRs."

Binding agents may be further capable of differentiating between patients with and without a cancer, such as colon cancer, using the representative assays provided herein. For example, antibodies or other binding agents that bind to a tumor protein will preferably generate a signal indicating the presence of a cancer in at least about 20% of patients with the disease, more preferably at least about 30% of patients. Alternatively, or in addition, the antibody will generate a negative signal indicating the absence of the disease in at least about 90% of individuals without the cancer. To determine whether a binding agent satisfies this requirement, biological samples (e.g., blood, sera, sputum, urine and/or tumor biopsies) from patients with and without a cancer (as determined using standard clinical tests) may be assayed as described herein for the presence of polypeptides that bind to the binding agent. Preferably, a statistically significant number of samples with and without the disease will be assayed. Each binding agent should satisfy the above criteria; however, those of ordinary skill in the art will recognize that binding agents may be used in combination to improve sensitivity.

Any agent that satisfies the above requirements may be a binding agent. For example, a binding agent may be a ribosome, with or without a peptide component, an RNA molecule or a polypeptide. In a preferred embodiment, a binding agent is an antibody or an antigen-binding fragment thereof. Antibodies may be prepared by any of a variety of techniques known to those of ordinary skill in the art. See, e.g., Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988. In general, antibodies can be produced by cell culture techniques, including the generation of monoclonal antibodies as described herein, or via transfection of antibody genes into suitable bacterial or mammalian cell hosts, in order to allow for the production of recombinant antibodies. In one technique, an immunogen comprising the polypeptide is initially injected into any of a wide variety of mammals (e.g., mice, rats, rabbits, sheep or goats). In this step, the polypeptides of this invention may serve as the immunogen without modification. Alternatively, particularly for relatively short polypeptides, a superior immune response may be elicited if the polypeptide is joined to a carrier protein, such as bovine serum albumin or keyhole limpet hemocyanin. The immunogen is injected into the animal host, preferably according to a predetermined schedule incorporating one or more booster immunizations, and the animals are bled periodically. Polyclonal antibodies specific for the polypeptide may then be purified from such antisera by, for example, affinity chromatography using the polypeptide coupled to a suitable solid support.

Monoclonal antibodies specific for an antigenic polypeptide of interest may be prepared, for example, using the technique of Kohler and Milstein, *Eur. J. Immunol.* 6:511-519, 1976, and improvements thereto. Briefly, these methods involve the preparation of immortal cell lines capable of producing antibodies having the desired specificity (i.e., reactivity with the polypeptide of interest). Such cell lines may be produced, for example, from spleen cells obtained from an animal immunized as described above. The spleen cells are then immortalized by, for example, fusion with a myeloma cell fusion partner, preferably one that is syngeneic with the immunized animal. A variety of fusion techniques may be employed. For example, the spleen cells and myeloma cells may be combined with a nonionic detergent for a few minutes and then plated at low density on a selective medium that supports the growth of hybrid

cells, but not myeloma cells. A preferred selection technique uses HAT (hypoxanthine, aminopterin, thymidine) selection. After a sufficient time, usually about 1 to 2 weeks, colonies of hybrids are observed. Single colonies are selected and their culture supernatants tested for binding activity against the polypeptide. Hybridomas having
5 high reactivity and specificity are preferred.

Monoclonal antibodies may be isolated from the supernatants of growing hybridoma colonies. In addition, various techniques may be employed to enhance the yield, such as injection of the hybridoma cell line into the peritoneal cavity of a suitable vertebrate host, such as a mouse. Monoclonal antibodies may then be harvested from
10 the ascites fluid or the blood. Contaminants may be removed from the antibodies by conventional techniques, such as chromatography, gel filtration, precipitation, and extraction. The polypeptides of this invention may be used in the purification process in, for example, an affinity chromatography step.

A number of therapeutically useful molecules are known in the art which
15 comprise antigen-binding sites that are capable of exhibiting immunological binding properties of an antibody molecule. The proteolytic enzyme papain preferentially cleaves IgG molecules to yield several fragments, two of which (the "F(ab)" fragments) each comprise a covalent heterodimer that includes an intact antigen-binding site. The enzyme pepsin is able to cleave IgG molecules to provide several fragments, including
20 the "F(ab')₂" fragment which comprises both antigen-binding sites. An "Fv" fragment can be produced by preferential proteolytic cleavage of an IgM, and on rare occasions IgG or IgA immunoglobulin molecule. Fv fragments are, however, more commonly derived using recombinant techniques known in the art. The Fv fragment includes a non-covalent V_H::V_L heterodimer including an antigen-binding site which retains much
25 of the antigen recognition and binding capabilities of the native antibody molecule. Inbar et al. (1972) Proc. Nat. Acad. Sci. USA 69:2659-2662; Hochman et al. (1976) Biochem 15:2706-2710; and Ehrlich et al. (1980) Biochem 19:4091-4096.

A single chain Fv ("sFv") polypeptide is a covalently linked V_H::V_L heterodimer which is expressed from a gene fusion including V_H- and V_L-encoding
30 genes linked by a peptide-encoding linker. Huston et al. (1988) Proc. Nat. Acad. Sci. USA 85(16):5879-5883. A number of methods have been described to discern chemical

structures for converting the naturally aggregated--but chemically separated--light and heavy polypeptide chains from an antibody V region into an sFv molecule which will fold into a three dimensional structure substantially similar to the structure of an antigen-binding site. See, *e.g.*, U.S. Pat. Nos. 5,091,513 and 5,132,405, to Huston et al.;
5 and U.S. Pat. No. 4,946,778, to Ladner et al.

Each of the above-described molecules includes a heavy chain and a light chain CDR set, respectively interposed between a heavy chain and a light chain FR set which provide support to the CDRs and define the spatial relationship of the CDRs relative to each other. As used herein, the term "CDR set" refers to the three
10 hypervariable regions of a heavy or light chain V region. Proceeding from the N-terminus of a heavy or light chain, these regions are denoted as "CDR1," "CDR2," and "CDR3" respectively. An antigen-binding site, therefore, includes six CDRs, comprising the CDR set from each of a heavy and a light chain V region. A polypeptide comprising a single CDR, (*e.g.*, a CDR1, CDR2 or CDR3) is referred to herein as a
15 "molecular recognition unit." Crystallographic analysis of a number of antigen-antibody complexes has demonstrated that the amino acid residues of CDRs form extensive contact with bound antigen, wherein the most extensive antigen contact is with the heavy chain CDR3. Thus, the molecular recognition units are primarily responsible for the specificity of an antigen-binding site.

20 As used herein, the term "FR set" refers to the four flanking amino acid sequences which frame the CDRs of a CDR set of a heavy or light chain V region. Some FR residues may contact bound antigen; however, FRs are primarily responsible for folding the V region into the antigen-binding site, particularly the FR residues directly adjacent to the CDRs. Within FRs, certain amino residues and certain structural
25 features are very highly conserved. In this regard, all V region sequences contain an internal disulfide loop of around 90 amino acid residues. When the V regions fold into a binding-site, the CDRs are displayed as projecting loop motifs which form an antigen-binding surface. It is generally recognized that there are conserved structural regions of FRs which influence the folded shape of the CDR loops into certain "canonical"
30 structures--regardless of the precise CDR amino acid sequence. Further, certain FR

residues are known to participate in non-covalent interdomain contacts which stabilize the interaction of the antibody heavy and light chains.

A number of "humanized" antibody molecules comprising an antigen-binding site derived from a non-human immunoglobulin have been described, including
5 chimeric antibodies having rodent V regions and their associated CDRs fused to human constant domains (Winter et al. (1991) Nature 349:293-299; Lobuglio et al. (1989) Proc. Nat. Acad. Sci. USA 86:4220-4224; Shaw et al. (1987) J Immunol. 138:4534-4538; and Brown et al. (1987) Cancer Res. 47:3577-3583), rodent CDRs grafted into a human supporting FR prior to fusion with an appropriate human antibody constant
10 domain (Riechmann et al. (1988) Nature 332:323-327; Verhoeyen et al. (1988) Science 239:1534-1536; and Jones et al. (1986) Nature 321:522-525), and rodent CDRs supported by recombinantly veneered rodent FRs (European Patent Publication No. 519,596, published Dec. 23, 1992). These "humanized" molecules are designed to minimize unwanted immunological response toward rodent antihuman antibody
15 molecules which limits the duration and effectiveness of therapeutic applications of those moieties in human recipients.

As used herein, the terms "veneered FRs" and "recombinantly veneered FRs" refer to the selective replacement of FR residues from, *e.g.*, a rodent heavy or light chain V region, with human FR residues in order to provide a xenogeneic molecule
20 comprising an antigen-binding site which retains substantially all of the native FR polypeptide folding structure. Veneering techniques are based on the understanding that the ligand binding characteristics of an antigen-binding site are determined primarily by the structure and relative disposition of the heavy and light chain CDR sets within the antigen-binding surface. Davies et al. (1990) Ann. Rev. Biochem. 59:439-473. Thus,
25 antigen binding specificity can be preserved in a humanized antibody only wherein the CDR structures, their interaction with each other, and their interaction with the rest of the V region domains are carefully maintained. By using veneering techniques, exterior (*e.g.*, solvent-accessible) FR residues which are readily encountered by the immune system are selectively replaced with human residues to provide a hybrid molecule that
30 comprises either a weakly immunogenic, or substantially non-immunogenic veneered surface.

The process of veneering makes use of the available sequence data for human antibody variable domains compiled by Kabat et al., in Sequences of Proteins of Immunological Interest, 4th ed., (U.S. Dept. of Health and Human Services, U.S. Government Printing Office, 1987), updates to the Kabat database, and other accessible U.S. and foreign databases (both nucleic acid and protein). Solvent accessibilities of V region amino acids can be deduced from the known three-dimensional structure for human and murine antibody fragments. There are two general steps in veneering a murine antigen-binding site. Initially, the FRs of the variable domains of an antibody molecule of interest are compared with corresponding FR sequences of human variable domains obtained from the above-identified sources. The most homologous human V regions are then compared residue by residue to corresponding murine amino acids. The residues in the murine FR which differ from the human counterpart are replaced by the residues present in the human moiety using recombinant techniques well known in the art. Residue switching is only carried out with moieties which are at least partially exposed (solvent accessible), and care is exercised in the replacement of amino acid residues which may have a significant effect on the tertiary structure of V region domains, such as proline, glycine and charged amino acids.

In this manner, the resultant "veneered" murine antigen-binding sites are thus designed to retain the murine CDR residues, the residues substantially adjacent to the CDRs, the residues identified as buried or mostly buried (solvent inaccessible), the residues believed to participate in non-covalent (e.g., electrostatic and hydrophobic) contacts between heavy and light chain domains, and the residues from conserved structural regions of the FRs which are believed to influence the "canonical" tertiary structures of the CDR loops. These design criteria are then used to prepare recombinant nucleotide sequences which combine the CDRs of both the heavy and light chain of a murine antigen-binding site into human-appearing FRs that can be used to transfect mammalian cells for the expression of recombinant human antibodies which exhibit the antigen specificity of the murine antibody molecule.

In another embodiment of the invention, monoclonal antibodies of the present invention may be coupled to one or more therapeutic agents. Suitable agents in this regard include radionuclides, differentiation inducers, drugs, toxins, and derivatives

thereof. Preferred radionuclides include ^{90}Y , ^{123}I , ^{125}I , ^{131}I , ^{186}Re , ^{188}Re , ^{211}At , and ^{212}Bi . Preferred drugs include methotrexate, and pyrimidine and purine analogs. Preferred differentiation inducers include phorbol esters and butyric acid. Preferred toxins include ricin, abrin, diphtheria toxin, cholera toxin, gelonin, *Pseudomonas* exotoxin, Shigella toxin, and pokeweed antiviral protein.

A therapeutic agent may be coupled (*e.g.*, covalently bonded) to a suitable monoclonal antibody either directly or indirectly (*e.g.*, via a linker group). A direct reaction between an agent and an antibody is possible when each possesses a substituent capable of reacting with the other. For example, a nucleophilic group, such as an amino or sulfhydryl group, on one may be capable of reacting with a carbonyl-containing group, such as an anhydride or an acid halide, or with an alkyl group containing a good leaving group (*e.g.*, a halide) on the other.

Alternatively, it may be desirable to couple a therapeutic agent and an antibody via a linker group. A linker group can function as a spacer to distance an antibody from an agent in order to avoid interference with binding capabilities. A linker group can also serve to increase the chemical reactivity of a substituent on an agent or an antibody, and thus increase the coupling efficiency. An increase in chemical reactivity may also facilitate the use of agents, or functional groups on agents, which otherwise would not be possible.

It will be evident to those skilled in the art that a variety of bifunctional or polyfunctional reagents, both homo- and hetero-functional (such as those described in the catalog of the Pierce Chemical Co., Rockford, IL), may be employed as the linker group. Coupling may be effected, for example, through amino groups, carboxyl groups, sulfhydryl groups or oxidized carbohydrate residues. There are numerous references describing such methodology, *e.g.*, U.S. Patent No. 4,671,958, to Rodwell et al.

Where a therapeutic agent is more potent when free from the antibody portion of the immunoconjugates of the present invention, it may be desirable to use a linker group which is cleavable during or upon internalization into a cell. A number of different cleavable linker groups have been described. The mechanisms for the intracellular release of an agent from these linker groups include cleavage by reduction of a disulfide bond (*e.g.*, U.S. Patent No. 4,489,710, to Spitler), by irradiation of a

photolabile bond (e.g., U.S. Patent No. 4,625,014, to Senter et al.), by hydrolysis of derivatized amino acid side chains (e.g., U.S. Patent No. 4,638,045, to Kohn et al.), by serum complement-mediated hydrolysis (e.g., U.S. Patent No. 4,671,958, to Rodwell et al.), and acid-catalyzed hydrolysis (e.g., U.S. Patent No. 4,569,789, to Blattler et al.).

5 It may be desirable to couple more than one agent to an antibody. In one embodiment, multiple molecules of an agent are coupled to one antibody molecule. In another embodiment, more than one type of agent may be coupled to one antibody. Regardless of the particular embodiment, immunoconjugates with more than one agent may be prepared in a variety of ways. For example, more than one agent may be
10 coupled directly to an antibody molecule, or linkers that provide multiple sites for attachment can be used. Alternatively, a carrier can be used.

A carrier may bear the agents in a variety of ways, including covalent bonding either directly or via a linker group. Suitable carriers include proteins such as albumins (e.g., U.S. Patent No. 4,507,234, to Kato et al.), peptides and polysaccharides
15 such as aminodextran (e.g., U.S. Patent No. 4,699,784, to Shih et al.). A carrier may also bear an agent by noncovalent bonding or by encapsulation, such as within a liposome vesicle (e.g., U.S. Patent Nos. 4,429,008 and 4,873,088). Carriers specific for radionuclide agents include radiohalogenated small molecules and chelating compounds. For example, U.S. Patent No. 4,735,792 discloses representative
20 radiohalogenated small molecules and their synthesis. A radionuclide chelate may be formed from chelating compounds that include those containing nitrogen and sulfur atoms as the donor atoms for binding the metal, or metal oxide, radionuclide. For example, U.S. Patent No. 4,673,562, to Davison et al. discloses representative chelating compounds and their synthesis.

25 T CELL COMPOSITIONS

The present invention, in another aspect, provides T cells specific for a tumor polypeptide disclosed herein, or for a variant or derivative thereof. Such cells may generally be prepared *in vitro* or *ex vivo*, using standard procedures. For example, T cells may be isolated from bone marrow, peripheral blood, or a fraction of bone
30 marrow or peripheral blood of a patient, using a commercially available cell separation

system, such as the Isolex™ System, available from Nexell Therapeutics, Inc. (Irvine, CA; see also U.S. Patent No. 5,240,856; U.S. Patent No. 5,215,926; WO 89/06280; WO 91/16116 and WO 92/07243). Alternatively, T cells may be derived from related or unrelated humans, non-human mammals, cell lines or cultures.

5 T cells may be stimulated with a polypeptide, polynucleotide encoding a polypeptide and/or an antigen presenting cell (APC) that expresses such a polypeptide. Such stimulation is performed under conditions and for a time sufficient to permit the generation of T cells that are specific for the polypeptide of interest. Preferably, a tumor polypeptide or polynucleotide of the invention is present within a delivery
10 vehicle, such as a microsphere, to facilitate the generation of specific T cells.

T cells are considered to be specific for a polypeptide of the present invention if the T cells specifically proliferate, secrete cytokines or kill target cells coated with the polypeptide or expressing a gene encoding the polypeptide. T cell specificity may be evaluated using any of a variety of standard techniques. For
15 example, within a chromium release assay or proliferation assay, a stimulation index of more than two fold increase in lysis and/or proliferation, compared to negative controls, indicates T cell specificity. Such assays may be performed, for example, as described in Chen et al., *Cancer Res.* 54:1065-1070, 1994. Alternatively, detection of the proliferation of T cells may be accomplished by a variety of known techniques. For
20 example, T cell proliferation can be detected by measuring an increased rate of DNA synthesis (e.g., by pulse-labeling cultures of T cells with tritiated thymidine and measuring the amount of tritiated thymidine incorporated into DNA). Contact with a tumor polypeptide (100 ng/ml - 100 µg/ml, preferably 200 ng/ml - 25 µg/ml) for 3 - 7 days will typically result in at least a two fold increase in proliferation of the T cells.
25 Contact as described above for 2-3 hours should result in activation of the T cells, as measured using standard cytokine assays in which a two fold increase in the level of cytokine release (e.g., TNF or IFN-γ) is indicative of T cell activation (see Coligan et al., *Current Protocols in Immunology*, vol. 1, Wiley Interscience (Greene 1998)). T cells that have been activated in response to a tumor polypeptide, polynucleotide or
30 polypeptide-expressing APC may be CD4⁺ and/or CD8⁺. Tumor polypeptide-specific T cells may be expanded using standard techniques. Within preferred embodiments, the T

cells are derived from a patient, a related donor or an unrelated donor, and are administered to the patient following stimulation and expansion.

For therapeutic purposes, $CD4^+$ or $CD8^+$ T cells that proliferate in response to a tumor polypeptide, polynucleotide or APC can be expanded in number either *in vitro* or *in vivo*. Proliferation of such T cells *in vitro* may be accomplished in a variety of ways. For example, the T cells can be re-exposed to a tumor polypeptide, or a short peptide corresponding to an immunogenic portion of such a polypeptide, with or without the addition of T cell growth factors, such as interleukin-2, and/or stimulator cells that synthesize a tumor polypeptide. Alternatively, one or more T cells that proliferate in the presence of the tumor polypeptide can be expanded in number by cloning. Methods for cloning cells are well known in the art, and include limiting dilution.

T CELL RECEPTOR COMPOSITIONS

The T cell receptor (TCR) consists of 2 different, highly variable polypeptide chains, termed the T-cell receptor α and β chains, that are linked by a disulfide bond (Janeway, Travers, Walport. *Immunobiology*. Fourth Ed., 148-159. Elsevier Science Ltd/Garland Publishing. 1999). The α/β heterodimer complexes with the invariant CD3 chains at the cell membrane. This complex recognizes specific antigenic peptides bound to MHC molecules. The enormous diversity of TCR specificities is generated much like immunoglobulin diversity, through somatic gene rearrangement. The β chain genes contain over 50 variable (V), 2 diversity (D), over 10 joining (J) segments, and 2 constant region segments (C). The α chain genes contain over 70 V segments, and over 60 J segments but no D segments, as well as one C segment. During T cell development in the thymus, the D to J gene rearrangement of the β chain occurs, followed by the V gene segment rearrangement to the DJ. This functional VDJ_β exon is transcribed and spliced to join to a C_β . For the α chain, a V_α gene segment rearranges to a J_α gene segment to create the functional exon that is then transcribed and spliced to the C_α . Diversity is further increased during the recombination process by the random addition of P and N-nucleotides between the V, D, and J segments of the β chain and between the V and J segments in the α chain

(Janeway, Travers, Walport. *Immunobiology*. Fourth Ed., 98 and 150. Elsevier Science Ltd/Garland Publishing. 1999).

The present invention, in another aspect, provides TCRs specific for a colon tumor polypeptide disclosed herein, or for a variant or derivative thereof. In accordance with the present invention, polynucleotide and amino acid sequences are provided for the V-J or V-D-J junctional regions or parts thereof for the alpha and beta chains of the T-cell receptor which recognize tumor polypeptides described herein. In general, this aspect of the invention relates to T-cell receptors which recognize or bind tumor polypeptides presented in the context of MHC. In a preferred embodiment the tumor antigens recognized by the T-cell receptors comprise a polypeptide of the present invention. For example, cDNA encoding a TCR specific for a colon tumor peptide can be isolated from T cells specific for a tumor polypeptide using standard molecular biological and recombinant DNA techniques.

This invention further includes the T-cell receptors or analogs thereof having substantially the same function or activity as the T-cell receptors of this invention which recognize or bind tumor polypeptides. Such receptors include, but are not limited to, a fragment of the receptor, or a substitution, addition or deletion mutant of a T-cell receptor provided herein. This invention also encompasses polypeptides or peptides that are substantially homologous to the T-cell receptors provided herein or that retain substantially the same activity. The term "analog" includes any protein or polypeptide having an amino acid residue sequence substantially identical to the T-cell receptors provided herein in which one or more residues, preferably no more than 5 residues, more preferably no more than 25 residues have been conservatively substituted with a functionally similar residue and which displays the functional aspects of the T-cell receptor as described herein.

The present invention further provides for suitable mammalian host cells, for example, non-specific T cells, that are transfected with a polynucleotide encoding TCRs specific for a polypeptide described herein, thereby rendering the host cell specific for the polypeptide. The α and β chains of the TCR may be contained on separate expression vectors or alternatively, on a single expression vector that also contains an internal ribosome entry site (IRES) for cap-independent translation of the gene downstream of the IRES. Said host cells expressing TCRs specific for the

polypeptide may be used, for example, for adoptive immunotherapy of colon cancer as discussed further below.

In further aspects of the present invention, cloned TCRs specific for a polypeptide recited herein may be used in a kit for the diagnosis of colon cancer. For example, the nucleic acid sequence or portions thereof, of colon tumor-specific TCRs can be used as probes or primers for the detection of expression of the rearranged genes encoding the specific TCR in a biological sample. Therefore, the present invention further provides for an assay for detecting messenger RNA or DNA encoding the TCR specific for a polypeptide.

10 PHARMACEUTICAL COMPOSITIONS

In additional embodiments, the present invention concerns formulation of one or more of the polynucleotide, polypeptide, T-cell, TCR, and/or antibody compositions disclosed herein in pharmaceutically-acceptable carriers for administration to a cell or an animal, either alone, or in combination with one or more other modalities of therapy.

It will be understood that, if desired, a composition as disclosed herein may be administered in combination with other agents as well, such as, *e.g.*, other proteins or polypeptides or various pharmaceutically-active agents. In fact, there is virtually no limit to other components that may also be included, given that the additional agents do not cause a significant adverse effect upon contact with the target cells or host tissues. The compositions may thus be delivered along with various other agents as required in the particular instance. Such compositions may be purified from host cells or other biological sources, or alternatively may be chemically synthesized as described herein. Likewise, such compositions may further comprise substituted or derivatized RNA or DNA compositions.

Therefore, in another aspect of the present invention, pharmaceutical compositions are provided comprising one or more of the polynucleotide, polypeptide, antibody, TCR, and/or T-cell compositions described herein in combination with a physiologically acceptable carrier. In certain preferred embodiments, the pharmaceutical compositions of the invention comprise immunogenic polynucleotide

and/or polypeptide compositions of the invention for use in prophylactic and therapeutic vaccine applications. Vaccine preparation is generally described in, for example, M.F. Powell and M.J. Newman, eds., "Vaccine Design (the subunit and adjuvant approach)," Plenum Press (NY, 1995). Generally, such compositions will comprise one or more
5 polynucleotide and/or polypeptide compositions of the present invention in combination with one or more immunostimulants.

It will be apparent that any of the pharmaceutical compositions described herein can contain pharmaceutically acceptable salts of the polynucleotides and polypeptides of the invention. Such salts can be prepared, for example, from
10 pharmaceutically acceptable non-toxic bases, including organic bases (e.g., salts of primary, secondary and tertiary amines and basic amino acids) and inorganic bases (e.g., sodium, potassium, lithium, ammonium, calcium and magnesium salts).

In another embodiment, illustrative immunogenic compositions, e.g., vaccine compositions, of the present invention comprise DNA encoding one or more of
15 the polypeptides as described above, such that the polypeptide is generated *in situ*. As noted above, the polynucleotide may be administered within any of a variety of delivery systems known to those of ordinary skill in the art. Indeed, numerous gene delivery techniques are well known in the art, such as those described by Rolland, *Crit. Rev. Therap. Drug Carrier Systems* 15:143-198, 1998, and references cited therein.
20 Appropriate polynucleotide expression systems will, of course, contain the necessary regulatory DNA regulatory sequences for expression in a patient (such as a suitable promoter and terminating signal). Alternatively, bacterial delivery systems may involve the administration of a bacterium (such as *Bacillus-Calmette-Guerrin*) that expresses an immunogenic portion of the polypeptide on its cell surface or secretes such an epitope.

25 Therefore, in certain embodiments, polynucleotides encoding immunogenic polypeptides described herein are introduced into suitable mammalian host cells for expression using any of a number of known viral-based systems. In one illustrative embodiment, retroviruses provide a convenient and effective platform for gene delivery systems. A selected nucleotide sequence encoding a polypeptide of the
30 present invention can be inserted into a vector and packaged in retroviral particles using techniques known in the art. The recombinant virus can then be isolated and delivered

to a subject. A number of illustrative retroviral systems have been described (*e.g.*, U.S. Pat. No. 5,219,740; Miller and Rosman (1989) *BioTechniques* 7:980-990; Miller, A. D. (1990) *Human Gene Therapy* 1:5-14; Scarpa et al. (1991) *Virology* 180:849-852; Burns et al. (1993) *Proc. Natl. Acad. Sci. USA* 90:8033-8037; and Boris-Lawrie and Temin (1993) *Cur. Opin. Genet. Develop.* 3:102-109.

In addition, a number of illustrative adenovirus-based systems have also been described. Unlike retroviruses which integrate into the host genome, adenoviruses persist extrachromosomally thus minimizing the risks associated with insertional mutagenesis (Haj-Ahmad and Graham (1986) *J. Virol.* 57:267-274; Bett et al. (1993) *J. Virol.* 67:5911-5921; Mittereder et al. (1994) *Human Gene Therapy* 5:717-729; Seth et al. (1994) *J. Virol.* 68:933-940; Barr et al. (1994) *Gene Therapy* 1:51-58; Berkner, K. L. (1988) *BioTechniques* 6:616-629; and Rich et al. (1993) *Human Gene Therapy* 4:461-476).

Various adeno-associated virus (AAV) vector systems have also been developed for polynucleotide delivery. AAV vectors can be readily constructed using techniques well known in the art. See, *e.g.*, U.S. Pat. Nos. 5,173,414 and 5,139,941; International Publication Nos. WO 92/01070 and WO 93/03769; Lebkowski et al. (1988) *Molec. Cell. Biol.* 8:3988-3996; Vincent et al. (1990) *Vaccines 90* (Cold Spring Harbor Laboratory Press); Carter, B. J. (1992) *Current Opinion in Biotechnology* 3:533-539; Muzyczka, N. (1992) *Current Topics in Microbiol. and Immunol.* 158:97-129; Kotin, R. M. (1994) *Human Gene Therapy* 5:793-801; Shelling and Smith (1994) *Gene Therapy* 1:165-169; and Zhou et al. (1994) *J. Exp. Med.* 179:1867-1875.

Additional viral vectors useful for delivering the polynucleotides encoding polypeptides of the present invention by gene transfer include those derived from the pox family of viruses, such as vaccinia virus and avian poxvirus. By way of example, vaccinia virus recombinants expressing the novel molecules can be constructed as follows. The DNA encoding a polypeptide is first inserted into an appropriate vector so that it is adjacent to a vaccinia promoter and flanking vaccinia DNA sequences, such as the sequence encoding thymidine kinase (TK). This vector is then used to transfect cells which are simultaneously infected with vaccinia. Homologous recombination serves to insert the vaccinia promoter plus the gene

encoding the polypeptide of interest into the viral genome. The resulting TK.sup.(-) recombinant can be selected by culturing the cells in the presence of 5-bromodeoxyuridine and picking viral plaques resistant thereto.

A vaccinia-based infection/transfection system can be conveniently used
5 to provide for inducible, transient expression or coexpression of one or more polypeptides described herein in host cells of an organism. In this particular system, cells are first infected in vitro with a vaccinia virus recombinant that encodes the bacteriophage T7 RNA polymerase. This polymerase displays exquisite specificity in that it only transcribes templates bearing T7 promoters. Following infection, cells are
10 transfected with the polynucleotide or polynucleotides of interest, driven by a T7 promoter. The polymerase expressed in the cytoplasm from the vaccinia virus recombinant transcribes the transfected DNA into RNA which is then translated into polypeptide by the host translational machinery. The method provides for high level, transient, cytoplasmic production of large quantities of RNA and its translation
15 products. See, e.g., Elroy-Stein and Moss, Proc. Natl. Acad. Sci. USA (1990) 87:6743-6747; Fuerst et al. Proc. Natl. Acad. Sci. USA (1986) 83:8122-8126.

Alternatively, avipoxviruses, such as the fowlpox and canarypox viruses, can also be used to deliver the coding sequences of interest. Recombinant avipox viruses, expressing immunogens from mammalian pathogens, are known to confer
20 protective immunity when administered to non-avian species. The use of an Avipox vector is particularly desirable in human and other mammalian species since members of the Avipox genus can only productively replicate in susceptible avian species and therefore are not infective in mammalian cells. Methods for producing recombinant Avipoxviruses are known in the art and employ genetic recombination, as described
25 above with respect to the production of vaccinia viruses. See, e.g., WO 91/12882; WO 89/03429; and WO 92/03545.

Any of a number of alphavirus vectors can also be used for delivery of polynucleotide compositions of the present invention, such as those vectors described in U.S. Patent Nos. 5,843,723; 6,015,686; 6,008,035 and 6,015,694. Certain vectors based
30 on Venezuelan Equine Encephalitis (VEE) can also be used, illustrative examples of which can be found in U.S. Patent Nos. 5,505,947 and 5,643,576.

Moreover, molecular conjugate vectors, such as the adenovirus chimeric vectors described in Michael et al. *J. Biol. Chem.* (1993) 268:6866-6869 and Wagner et al. *Proc. Natl. Acad. Sci. USA* (1992) 89:6099-6103, can also be used for gene delivery under the invention.

5 Additional illustrative information on these and other known viral-based delivery systems can be found, for example, in Fisher-Hoch et al., *Proc. Natl. Acad. Sci. USA* 86:317-321, 1989; Flexner et al., *Ann. N.Y. Acad. Sci.* 569:86-103, 1989; Flexner et al., *Vaccine* 8:17-21, 1990; U.S. Patent Nos. 4,603,112, 4,769,330, and 5,017,487; WO 89/01973; U.S. Patent No. 4,777,127; GB 2,200,651; EP 0,345,242; WO 91/02805;
10 Berkner, *Biotechniques* 6:616-627, 1988; Rosenfeld et al., *Science* 252:431-434, 1991; Kolls et al., *Proc. Natl. Acad. Sci. USA* 91:215-219, 1994; Kass-Eisler et al., *Proc. Natl. Acad. Sci. USA* 90:11498-11502, 1993; Guzman et al., *Circulation* 88:2838-2848, 1993; and Guzman et al., *Cir. Res.* 73:1202-1207, 1993.

In certain embodiments, a polynucleotide may be integrated into the
15 genome of a target cell. This integration may be in the specific location and orientation via homologous recombination (gene replacement) or it may be integrated in a random, non-specific location (gene augmentation). In yet further embodiments, the polynucleotide may be stably maintained in the cell as a separate, episomal segment of DNA. Such polynucleotide segments or "episomes" encode sequences sufficient to
20 permit maintenance and replication independent of or in synchronization with the host cell cycle. The manner in which the expression construct is delivered to a cell and where in the cell the polynucleotide remains is dependent on the type of expression construct employed.

In another embodiment of the invention, a polynucleotide is
25 administered/delivered as "naked" DNA, for example as described in Ulmer et al., *Science* 259:1745-1749, 1993 and reviewed by Cohen, *Science* 259:1691-1692, 1993. The uptake of naked DNA may be increased by coating the DNA onto biodegradable beads, which are efficiently transported into the cells.

In still another embodiment, a composition of the present invention can
30 be delivered via a particle bombardment approach, many of which have been described. In one illustrative example, gas-driven particle acceleration can be achieved with

devices such as those manufactured by Powderject Pharmaceuticals PLC (Oxford, UK) and Powderject Vaccines Inc. (Madison, WI), some examples of which are described in U.S. Patent Nos. 5,846,796; 6,010,478; 5,865,796; 5,584,807; and EP Patent No. 0500 799. This approach offers a needle-free delivery approach wherein a dry powder
5 formulation of microscopic particles, such as polynucleotide or polypeptide particles, are accelerated to high speed within a helium gas jet generated by a hand held device, propelling the particles into a target tissue of interest.

In a related embodiment, other devices and methods that may be useful for gas-driven needle-less injection of compositions of the present invention include
10 those provided by Bioject, Inc. (Portland, OR), some examples of which are described in U.S. Patent Nos. 4,790,824; 5,064,413; 5,312,335; 5,383,851; 5,399,163; 5,520,639 and 5,993,412.

According to another embodiment, the pharmaceutical compositions described herein will comprise one or more immunostimulants in addition to the
15 immunogenic polynucleotide, polypeptide, antibody, T-cell, TCR, and/or APC compositions of this invention. An immunostimulant refers to essentially any substance that enhances or potentiates an immune response (antibody and/or cell-mediated) to an exogenous antigen. One preferred type of immunostimulant comprises an adjuvant. Many adjuvants contain a substance designed to protect the antigen from rapid
20 catabolism, such as aluminum hydroxide or mineral oil, and a stimulator of immune responses, such as lipid A, *Bordetella pertussis* or *Mycobacterium tuberculosis* derived proteins. Certain adjuvants are commercially available as, for example, Freund's Incomplete Adjuvant and Complete Adjuvant (Difco Laboratories, Detroit, MI); Merck Adjuvant 65 (Merck and Company, Inc., Rahway, NJ); AS-2 (SmithKline Beecham,
25 Philadelphia, PA); aluminum salts such as aluminum hydroxide gel (alum) or aluminum phosphate; salts of calcium, iron or zinc; an insoluble suspension of acylated tyrosine; acylated sugars; cationically or anionically derivatized polysaccharides; polyphosphazenes; biodegradable microspheres; monophosphoryl lipid A and quil A. Cytokines, such as GM-CSF, interleukin-2, -7, -12, and other like growth factors, may
30 also be used as adjuvants.

Within certain embodiments of the invention, the adjuvant composition is preferably one that induces an immune response predominantly of the Th1 type. High levels of Th1-type cytokines (*e.g.*, IFN- γ , TNF α , IL-2 and IL-12) tend to favor the induction of cell mediated immune responses to an administered antigen. In contrast, 5 high levels of Th2-type cytokines (*e.g.*, IL-4, IL-5, IL-6 and IL-10) tend to favor the induction of humoral immune responses. Following application of a vaccine as provided herein, a patient will support an immune response that includes Th1- and Th2-type responses. Within a preferred embodiment, in which a response is predominantly Th1-type, the level of Th1-type cytokines will increase to a greater extent than the level 10 of Th2-type cytokines. The levels of these cytokines may be readily assessed using standard assays. For a review of the families of cytokines, see Mosmann and Coffman, *Ann. Rev. Immunol.* 7:145-173, 1989.

Certain preferred adjuvants for eliciting a predominantly Th1-type response include, for example, a combination of monophosphoryl lipid A, preferably 3- 15 de-O-acylated monophosphoryl lipid A, together with an aluminum salt. MPL[®] adjuvants are available from Corixa Corporation (Seattle, WA; *see*, for example, US Patent Nos. 4,436,727; 4,877,611; 4,866,034 and 4,912,094). CpG-containing oligonucleotides (in which the CpG dinucleotide is unmethylated) also induce a predominantly Th1 response. Such oligonucleotides are well known and are described, 20 for example, in WO 96/02555, WO 99/33488 and U.S. Patent Nos. 6,008,200 and 5,856,462. Immunostimulatory DNA sequences are also described, for example, by Sato et al., *Science* 273:352, 1996. Another preferred adjuvant comprises a saponin, such as Quil A, or derivatives thereof, including QS21 and QS7 (Aquila Biopharmaceuticals Inc., Framingham, MA); Escin; Digitonin; or *Gypsophila* or 25 *Chenopodium quinoa* saponins. Other preferred formulations include more than one saponin in the adjuvant combinations of the present invention, for example combinations of at least two of the following group comprising QS21, QS7, Quil A, β -escin, or digitonin.

Alternatively the saponin formulations may be combined with vaccine 30 vehicles composed of chitosan or other polycationic polymers, polylactide and polylactide-co-glycolide particles, poly-N-acetyl glucosamine-based polymer matrix,

particles composed of polysaccharides or chemically modified polysaccharides, liposomes and lipid-based particles, particles composed of glycerol monoesters, etc. The saponins may also be formulated in the presence of cholesterol to form particulate structures such as liposomes or ISCOMs. Furthermore, the saponins may be formulated
5 together with a polyoxyethylene ether or ester, in either a non-particulate solution or suspension, or in a particulate structure such as a paucilamellar liposome or ISCOM. The saponins may also be formulated with excipients such as Carbopol^R to increase viscosity, or may be formulated in a dry powder form with a powder excipient such as lactose.

10 In one preferred embodiment, the adjuvant system includes the combination of a monophosphoryl lipid A and a saponin derivative, such as the combination of QS21 and 3D-MPL[®] adjuvant, as described in WO 94/00153, or a less reactogenic composition where the QS21 is quenched with cholesterol, as described in WO 96/33739. Other preferred formulations comprise an oil-in-water emulsion and
15 tocopherol. Another particularly preferred adjuvant formulation employing QS21, 3D-MPL[®] adjuvant and tocopherol in an oil-in-water emulsion is described in WO 95/17210.

Another enhanced adjuvant system involves the combination of a CpG-containing oligonucleotide and a saponin derivative particularly the combination of
20 CpG and QS21 is disclosed in WO 00/09159. Preferably the formulation additionally comprises an oil in water emulsion and tocopherol.

Additional illustrative adjuvants for use in the pharmaceutical compositions of the invention include Montanide ISA 720 (Seppic, France), SAF (Chiron, California, United States), ISCOMS (CSL), MF-59 (Chiron), the SBAS series
25 of adjuvants (e.g., SBAS-2 or SBAS-4, available from SmithKline Beecham, Rixensart, Belgium), Detox (Enhanzyn[®]) (Corixa, Hamilton, MT), RC-529 (Corixa, Hamilton, MT) and other aminoalkyl glucosaminide 4-phosphates (AGPs), such as those described in pending U.S. Patent Application Serial Nos. 08/853,826 and 09/074,720, the disclosures of which are incorporated herein by reference in their entireties, and
30 polyoxyethylene ether adjuvants such as those described in WO 99/52549A1.

Other preferred adjuvants include adjuvant molecules of the general formula



wherein, n is 1-50, A is a bond or $-\text{C}(\text{O})-$, R is C_{1-50} alkyl or Phenyl C_{1-50} alkyl.

5 One embodiment of the present invention consists of a vaccine formulation comprising a polyoxyethylene ether of general formula (I), wherein n is between 1 and 50, preferably 4-24, most preferably 9; the R component is C_{1-50} , preferably $\text{C}_4\text{-C}_{20}$ alkyl and most preferably C_{12} alkyl, and A is a bond. The concentration of the polyoxyethylene ethers should be in the range 0.1-20%, preferably
10 from 0.1-10%, and most preferably in the range 0.1-1%. Preferred polyoxyethylene ethers are selected from the following group: polyoxyethylene-9-lauryl ether, polyoxyethylene-9-stearyl ether, polyoxyethylene-8-stearyl ether, polyoxyethylene-4-lauryl ether, polyoxyethylene-35-lauryl ether, and polyoxyethylene-23-lauryl ether. Polyoxyethylene ethers such as polyoxyethylene lauryl ether are described in the Merck
15 index (12th edition: entry 7717). These adjuvant molecules are described in WO 99/52549.

The polyoxyethylene ether according to the general formula (I) above may, if desired, be combined with another adjuvant. For example, a preferred adjuvant combination is preferably with CpG as described in the pending UK patent application
20 GB 9820956.2.

According to another embodiment of this invention, an immunogenic composition described herein is delivered to a host via antigen presenting cells (APCs), such as dendritic cells, macrophages, B cells, monocytes and other cells that may be engineered to be efficient APCs. Such cells may, but need not, be genetically modified
25 to increase the capacity for presenting the antigen, to improve activation and/or maintenance of the T cell response, to have anti-tumor effects *per se* and/or to be immunologically compatible with the receiver (*i.e.*, matched HLA haplotype). APCs may generally be isolated from any of a variety of biological fluids and organs, including tumor and peritumoral tissues, and may be autologous, allogeneic, syngeneic
30 or xenogeneic cells.

Certain preferred embodiments of the present invention use dendritic cells or progenitors thereof as antigen-presenting cells. Dendritic cells are highly potent APCs (Banchereau and Steinman, *Nature* 392:245-251, 1998) and have been shown to be effective as a physiological adjuvant for eliciting prophylactic or therapeutic antitumor immunity (see Timmerman and Levy, *Ann. Rev. Med.* 50:507-529, 1999). In general, dendritic cells may be identified based on their typical shape (stellate *in situ*, with marked cytoplasmic processes (dendrites) visible *in vitro*), their ability to take up, process and present antigens with high efficiency and their ability to activate naïve T cell responses. Dendritic cells may, of course, be engineered to express specific cell-surface receptors or ligands that are not commonly found on dendritic cells *in vivo* or *ex vivo*, and such modified dendritic cells are contemplated by the present invention. As an alternative to dendritic cells, secreted vesicles antigen-loaded dendritic cells (called exosomes) may be used within a vaccine (see Zitvogel et al., *Nature Med.* 4:594-600, 1998).

Dendritic cells and progenitors may be obtained from peripheral blood, bone marrow, tumor-infiltrating cells, peritumoral tissues-infiltrating cells, lymph nodes, spleen, skin, umbilical cord blood or any other suitable tissue or fluid. For example, dendritic cells may be differentiated *ex vivo* by adding a combination of cytokines such as GM-CSF, IL-4, IL-13 and/or TNF α to cultures of monocytes harvested from peripheral blood. Alternatively, CD34 positive cells harvested from peripheral blood, umbilical cord blood or bone marrow may be differentiated into dendritic cells by adding to the culture medium combinations of GM-CSF, IL-3, TNF α , CD40 ligand, LPS, flt3 ligand and/or other compound(s) that induce differentiation, maturation and proliferation of dendritic cells.

Dendritic cells are conveniently categorized as "immature" and "mature" cells, which allows a simple way to discriminate between two well characterized phenotypes. However, this nomenclature should not be construed to exclude all possible intermediate stages of differentiation. Immature dendritic cells are characterized as APC with a high capacity for antigen uptake and processing, which correlates with the high expression of Fc γ receptor and mannose receptor. The mature phenotype is typically characterized by a lower expression of these markers, but a high

expression of cell surface molecules responsible for T cell activation such as class I and class II MHC, adhesion molecules (e.g., CD54 and CD11) and costimulatory molecules (e.g., CD40, CD80, CD86 and 4-1BB).

APCs may generally be transfected with a polynucleotide of the invention (or portion or other variant thereof) such that the encoded polypeptide, or an immunogenic portion thereof, is expressed on the cell surface. Such transfection may take place *ex vivo*, and a pharmaceutical composition comprising such transfected cells may then be used for therapeutic purposes, as described herein. Alternatively, a gene delivery vehicle that targets a dendritic or other antigen presenting cell may be administered to a patient, resulting in transfection that occurs *in vivo*. *In vivo* and *ex vivo* transfection of dendritic cells, for example, may generally be performed using any methods known in the art, such as those described in WO 97/24447, or the gene gun approach described by Mahvi et al., *Immunology and cell Biology* 75:456-460, 1997. Antigen loading of dendritic cells may be achieved by incubating dendritic cells or progenitor cells with the tumor polypeptide, DNA (naked or within a plasmid vector) or RNA; or with antigen-expressing recombinant bacterium or viruses (e.g., vaccinia, fowlpox, adenovirus or lentivirus vectors). Prior to loading, the polypeptide may be covalently conjugated to an immunological partner that provides T cell help (e.g., a carrier molecule). Alternatively, a dendritic cell may be pulsed with a non-conjugated immunological partner, separately or in the presence of the polypeptide.

While any suitable carrier known to those of ordinary skill in the art may be employed in the pharmaceutical compositions of this invention, the type of carrier will typically vary depending on the mode of administration. Compositions of the present invention may be formulated for any appropriate manner of administration, including for example, topical, oral, nasal, mucosal, intravenous, intracranial, intraperitoneal, subcutaneous and intramuscular administration.

Carriers for use within such pharmaceutical compositions are biocompatible, and may also be biodegradable. In certain embodiments, the formulation preferably provides a relatively constant level of active component release. In other embodiments, however, a more rapid rate of release immediately upon administration may be desired. The formulation of such compositions is well within the

level of ordinary skill in the art using known techniques. Illustrative carriers useful in this regard include microparticles of poly(lactide-co-glycolide), polyacrylate, latex, starch, cellulose, dextran and the like. Other illustrative delayed-release carriers include supramolecular biovectors, which comprise a non-liquid hydrophilic core (e.g.,
5 a cross-linked polysaccharide or oligosaccharide) and, optionally, an external layer comprising an amphiphilic compound, such as a phospholipid (*see e.g.*, U.S. Patent No. 5,151,254 and PCT applications WO 94/20078, WO/94/23701 and WO 96/06638). The amount of active compound contained within a sustained release formulation depends upon the site of implantation, the rate and expected duration of release and the nature of
10 the condition to be treated or prevented.

In another illustrative embodiment, biodegradable microspheres (e.g., polylactate polyglycolate) are employed as carriers for the compositions of this invention. Suitable biodegradable microspheres are disclosed, for example, in U.S. Patent Nos. 4,897,268; 5,075,109; 5,928,647; 5,811,128; 5,820,883; 5,853,763;
15 5,814,344, 5,407,609 and 5,942,252. Modified hepatitis B core protein carrier systems, such as described in WO/99 40934, and references cited therein, will also be useful for many applications. Another illustrative carrier/delivery system employs a carrier comprising particulate-protein complexes, such as those described in U.S. Patent No. 5,928,647, which are capable of inducing a class I-restricted cytotoxic T lymphocyte
20 responses in a host.

In another illustrative embodiment, calcium phosphate core particles are employed as carriers, vaccine adjuvants, or as controlled release matrices for the compositions of this invention. Exemplary calcium phosphate particles are disclosed, for example, in published patent application No. WO/0046147.

25 The pharmaceutical compositions of the invention will often further comprise one or more buffers (e.g., neutral buffered saline or phosphate buffered saline), carbohydrates (e.g., glucose, mannose, sucrose or dextrans), mannitol, proteins, polypeptides or amino acids such as glycine, antioxidants, bacteriostats, chelating agents such as EDTA or glutathione, adjuvants (e.g., aluminum hydroxide), solutes that
30 render the formulation isotonic, hypotonic or weakly hypertonic with the blood of a

recipient, suspending agents, thickening agents and/or preservatives. Alternatively, compositions of the present invention may be formulated as a lyophilizate.

The pharmaceutical compositions described herein may be presented in unit-dose or multi-dose containers, such as sealed ampoules or vials. Such containers
5 are typically sealed in such a way to preserve the sterility and stability of the formulation until use. In general, formulations may be stored as suspensions, solutions or emulsions in oily or aqueous vehicles. Alternatively, a pharmaceutical composition may be stored in a freeze-dried condition requiring only the addition of a sterile liquid carrier immediately prior to use.

10 The development of suitable dosing and treatment regimens for using the particular compositions described herein in a variety of treatment regimens, including *e.g.*, oral, parenteral, intravenous, intranasal, and intramuscular administration and formulation, is well known in the art, some of which are briefly discussed below for general purposes of illustration.

15 In certain applications, the pharmaceutical compositions disclosed herein may be delivered *via* oral administration to an animal. As such, these compositions may be formulated with an inert diluent or with an assimilable edible carrier, or they may be enclosed in hard- or soft-shell gelatin capsule, or they may be compressed into tablets, or they may be incorporated directly with the food of the diet.

20 The active compounds may even be incorporated with excipients and used in the form of ingestible tablets, buccal tables, troches, capsules, elixirs, suspensions, syrups, wafers, and the like (see, for example, Mathiowitz *et al.*, Nature 1997 Mar 27;386(6623):410-4; Hwang *et al.*, Crit Rev Ther Drug Carrier Syst 1998;15(3):243-84; U. S. Patent 5,641,515; U. S. Patent 5,580,579 and U. S. Patent
25 5,792,451). Tablets, troches, pills, capsules and the like may also contain any of a variety of additional components, for example, a binder, such as gum tragacanth, acacia, cornstarch, or gelatin; excipients, such as dicalcium phosphate; a disintegrating agent, such as corn starch, potato starch, alginic acid and the like; a lubricant, such as magnesium stearate; and a sweetening agent, such as sucrose, lactose or saccharin may
30 be added or a flavoring agent, such as peppermint, oil of wintergreen, or cherry flavoring. When the dosage unit form is a capsule, it may contain, in addition to

materials of the above type, a liquid carrier. Various other materials may be present as coatings or to otherwise modify the physical form of the dosage unit. For instance, tablets, pills, or capsules may be coated with shellac, sugar, or both. Of course, any material used in preparing any dosage unit form should be pharmaceutically pure and substantially non-toxic in the amounts employed. In addition, the active compounds may be incorporated into sustained-release preparation and formulations.

Typically, these formulations will contain at least about 0.1% of the active compound or more, although the percentage of the active ingredient(s) may, of course, be varied and may conveniently be between about 1 or 2% and about 60% or 70% or more of the weight or volume of the total formulation. Naturally, the amount of active compound(s) in each therapeutically useful composition may be prepared in such a way that a suitable dosage will be obtained in any given unit dose of the compound. Factors such as solubility, bioavailability, biological half-life, route of administration, product shelf life, as well as other pharmacological considerations will be contemplated by one skilled in the art of preparing such pharmaceutical formulations, and as such, a variety of dosages and treatment regimens may be desirable.

For oral administration the compositions of the present invention may alternatively be incorporated with one or more excipients in the form of a mouthwash, dentifrice, buccal tablet, oral spray, or sublingual orally-administered formulation. Alternatively, the active ingredient may be incorporated into an oral solution such as one containing sodium borate, glycerin and potassium bicarbonate, or dispersed in a dentifrice, or added in a therapeutically-effective amount to a composition that may include water, binders, abrasives, flavoring agents, foaming agents, and humectants. Alternatively the compositions may be fashioned into a tablet or solution form that may be placed under the tongue or otherwise dissolved in the mouth.

In certain circumstances it will be desirable to deliver the pharmaceutical compositions disclosed herein parenterally, intravenously, intramuscularly, or even intraperitoneally. Such approaches are well known to the skilled artisan, some of which are further described, for example, in U. S. Patent 5,543,158; U. S. Patent 5,641,515 and U. S. Patent 5,399,363. In certain embodiments, solutions of the active compounds as free base or pharmacologically acceptable salts may be prepared in water suitably

mixed with a surfactant, such as hydroxypropylcellulose. Dispersions may also be prepared in glycerol, liquid polyethylene glycols, and mixtures thereof and in oils. Under ordinary conditions of storage and use, these preparations generally will contain a preservative to prevent the growth of microorganisms.

- 5 Illustrative pharmaceutical forms suitable for injectable use include sterile aqueous solutions or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersions (for example, see U. S. Patent 5,466,468). In all cases the form must be sterile and must be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and
10 storage and must be preserved against the contaminating action of microorganisms, such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (*e.g.*, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), suitable mixtures thereof, and/or vegetable oils. Proper fluidity may be maintained, for example, by the use of a coating, such as
15 lecithin, by the maintenance of the required particle size in the case of dispersion and/or by the use of surfactants. The prevention of the action of microorganisms can be facilitated by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, sorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars or sodium chloride.
20 Prolonged absorption of the injectable compositions can be brought about by the use in the compositions of agents delaying absorption, for example, aluminum monostearate and gelatin.

In one embodiment, for parenteral administration in an aqueous solution, the solution should be suitably buffered if necessary and the liquid diluent first rendered
25 isotonic with sufficient saline or glucose. These particular aqueous solutions are especially suitable for intravenous, intramuscular, subcutaneous and intraperitoneal administration. In this connection, a sterile aqueous medium that can be employed will be known to those of skill in the art in light of the present disclosure. For example, one dosage may be dissolved in 1 ml of isotonic NaCl solution and either added to 1000 ml
30 of hypodermoclysis fluid or injected at the proposed site of infusion, (see for example, "Remington's Pharmaceutical Sciences" 15th Edition, pages 1035-1038 and 1570-

1580). Some variation in dosage will necessarily occur depending on the condition of the subject being treated. Moreover, for human administration, preparations will of course preferably meet sterility, pyrogenicity, and the general safety and purity standards as required by FDA Office of Biologics standards.

5 In another embodiment of the invention, the compositions disclosed herein may be formulated in a neutral or salt form. Illustrative pharmaceutically-acceptable salts include the acid addition salts (formed with the free amino groups of the protein) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids as acetic, oxalic, 10 tartaric, mandelic, and the like. Salts formed with the free carboxyl groups can also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, histidine, procaine and the like. Upon formulation, solutions will be administered in a manner compatible with the dosage formulation and in such amount 15 as is therapeutically effective.

The carriers can further comprise any and all solvents, dispersion media, vehicles, coatings, diluents, antibacterial and antifungal agents, isotonic and absorption delaying agents, buffers, carrier solutions, suspensions, colloids, and the like. The use of such media and agents for pharmaceutical active substances is well known in the art. 20 Except insofar as any conventional media or agent is incompatible with the active ingredient, its use in the therapeutic compositions is contemplated. Supplementary active ingredients can also be incorporated into the compositions. The phrase "pharmaceutically-acceptable" refers to molecular entities and compositions that do not produce an allergic or similar untoward reaction when administered to a human.

25 In certain embodiments, the pharmaceutical compositions may be delivered by intranasal sprays, inhalation, and/or other aerosol delivery vehicles. Methods for delivering genes, nucleic acids, and peptide compositions directly to the lungs *via* nasal aerosol sprays has been described, *e.g.*, in U. S. Patent 5,756,353 and U. S. Patent 5,804,212. Likewise, the delivery of drugs using intranasal microparticle 30 resins (Takenaga *et al.*, J Controlled Release 1998 Mar 2;52(1-2):81-7) and lysophosphatidyl-glycerol compounds (U. S. Patent 5,725,871) are also well-known in

the pharmaceutical arts. Likewise, illustrative transmucosal drug delivery in the form of a polytetrafluoroethylene support matrix is described in U. S. Patent 5,780,045.

In certain embodiments, liposomes, nanocapsules, microparticles, lipid particles, vesicles, and the like, are used for the introduction of the compositions of the present invention into suitable host cells/organisms. In particular, the compositions of the present invention may be formulated for delivery either encapsulated in a lipid particle, a liposome, a vesicle, a nanosphere, or a nanoparticle or the like. Alternatively, compositions of the present invention can be bound, either covalently or non-covalently, to the surface of such carrier vehicles.

The formation and use of liposome and liposome-like preparations as potential drug carriers is generally known to those of skill in the art (see for example, Lasic, Trends Biotechnol 1998 Jul;16(7):307-21; Takakura, Nippon Rinsho 1998 Mar;56(3):691-5; Chandran *et al.*, Indian J Exp Biol. 1997 Aug;35(8):801-9; Margalit, Crit Rev Ther Drug Carrier Syst. 1995;12(2-3):233-61; U.S. Patent 5,567,434; U.S. Patent 5,552,157; U.S. Patent 5,565,213; U.S. Patent 5,738,868 and U.S. Patent 5,795,587, each specifically incorporated herein by reference in its entirety).

Liposomes have been used successfully with a number of cell types that are normally difficult to transfect by other procedures, including T cell suspensions, primary hepatocyte cultures and PC12 cells (Renneisen *et al.*, J Biol Chem. 1990 Sep 25;265(27):16337-42; Muller *et al.*, DNA Cell Biol. 1990 Apr;9(3):221-9). In addition, liposomes are free of the DNA length constraints that are typical of viral-based delivery systems. Liposomes have been used effectively to introduce genes, various drugs, radiotherapeutic agents, enzymes, viruses, transcription factors, allosteric effectors and the like, into a variety of cultured cell lines and animals. Furthermore, the use of liposomes does not appear to be associated with autoimmune responses or unacceptable toxicity after systemic delivery.

In certain embodiments, liposomes are formed from phospholipids that are dispersed in an aqueous medium and spontaneously form multilamellar concentric bilayer vesicles (also termed multilamellar vesicles (MLVs)).

Alternatively, in other embodiments, the invention provides for pharmaceutically-acceptable nanocapsule formulations of the compositions of the

present invention. Nanocapsules can generally entrap compounds in a stable and reproducible way (see, for example, Quintanar-Guerrero *et al.*, Drug Dev Ind Pharm. 1998 Dec;24(12):1113-28). To avoid side effects due to intracellular polymeric overloading, such ultrafine particles (sized around 0.1 μm) may be designed using
5 polymers able to be degraded *in vivo*. Such particles can be made as described, for example, by Couvreur *et al.*, Crit Rev Ther Drug Carrier Syst. 1988;5(1):1-20; zur Muhlen *et al.*, Eur J Pharm Biopharm. 1998 Mar;45(2):149-55; Zambaux *et al.* J Controlled Release. 1998 Jan 2;50(1-3):31-40; and U. S. Patent 5,145,684.

CANCER THERAPEUTIC METHODS

10 Immunologic approaches to cancer therapy are based on the recognition that cancer cells can often evade the body's defenses against aberrant or foreign cells and molecules, and that these defenses might be therapeutically stimulated to regain the lost ground, *e.g.* pgs. 623-648 in Klein, Immunology (Wiley-Interscience, New York, 1982). Numerous recent observations that various immune effectors can directly or
15 indirectly inhibit growth of tumors has led to renewed interest in this approach to cancer therapy, *e.g.* Jager, *et al.*, Oncology 2001;60(1):1-7; Renner, *et al.*, Ann Hematol 2000 Dec;79(12):651-9.

Four basic cell types whose function has been associated with antitumor cell immunity and the elimination of tumor cells from the body are: i) B-lymphocytes
20 which secrete immunoglobulins into the blood plasma for identifying and labeling the nonself invader cells; ii) monocytes which secrete the complement proteins that are responsible for lysing and processing the immunoglobulin-coated target invader cells; iii) natural killer lymphocytes having two mechanisms for the destruction of tumor cells, antibody-dependent cellular cytotoxicity and natural killing; and iv) T-
25 lymphocytes possessing antigen-specific receptors and having the capacity to recognize a tumor cell carrying complementary marker molecules (Schreiber, H., 1989, in Fundamental Immunology (ed). W. E. Paul, pp. 923-955).

Cancer immunotherapy generally focuses on inducing humoral immune responses, cellular immune responses, or both. Moreover, it is well established that
30 induction of CD4⁺ T helper cells is necessary in order to secondarily induce either

antibodies or cytotoxic CD8⁺ T cells. Polypeptide antigens that are selective or ideally specific for cancer cells, particularly colon cancer cells, offer a powerful approach for inducing immune responses against colon cancer, and are an important aspect of the present invention.

5 Therefore, in further aspects of the present invention, the pharmaceutical compositions described herein may be used to stimulate an immune response against cancer, particularly for the immunotherapy of colon cancer. Within such methods, the pharmaceutical compositions described herein are administered to a patient, typically a warm-blooded animal, preferably a human. A patient may or may not be afflicted with
10 cancer. Pharmaceutical compositions and vaccines may be administered either prior to or following surgical removal of primary tumors and/or treatment such as administration of radiotherapy or conventional chemotherapeutic drugs. As discussed above, administration of the pharmaceutical compositions may be by any suitable method, including administration by intravenous, intraperitoneal, intramuscular,
15 subcutaneous, intranasal, intradermal, anal, vaginal, topical and oral routes.

 Within certain embodiments, immunotherapy may be active immunotherapy, in which treatment relies on the *in vivo* stimulation of the endogenous host immune system to react against tumors with the administration of immune response-modifying agents (such as polypeptides and polynucleotides as provided
20 herein).

 Within other embodiments, immunotherapy may be passive immunotherapy, in which treatment involves the delivery of agents with established tumor-immune reactivity (such as effector cells or antibodies) that can directly or indirectly mediate antitumor effects and does not necessarily depend on an intact host
25 immune system. Examples of effector cells include T cells as discussed above, T lymphocytes (such as CD8⁺ cytotoxic T lymphocytes and CD4⁺ T-helper tumor-infiltrating lymphocytes), killer cells (such as Natural Killer cells and lymphokine-activated killer cells), B cells and antigen-presenting cells (such as dendritic cells and macrophages) expressing a polypeptide provided herein. T cell receptors and antibody
30 receptors specific for the polypeptides recited herein may be cloned, expressed and transferred into other vectors or effector cells for adoptive immunotherapy. The

polypeptides provided herein may also be used to generate antibodies or anti-idiotypic antibodies (as described above and in U.S. Patent No. 4,918,164) for passive immunotherapy.

Monoclonal antibodies may be labeled with any of a variety of labels for
5 desired selective usages in detection, diagnostic assays or therapeutic applications (as described in U.S. Patent Nos. 6,090,365; 6,015,542; 5,843,398; 5,595,721; and 4,708,930, hereby incorporated by reference in their entirety as if each was incorporated individually). In each case, the binding of the labelled monoclonal antibody to the
10 determinant site of the antigen will signal detection or delivery of a particular therapeutic agent to the antigenic determinant on the non-normal cell. A further object of this invention is to provide the specific monoclonal antibody suitably labelled for achieving such desired selective usages thereof.

Effector cells may generally be obtained in sufficient quantities for adoptive immunotherapy by growth *in vitro*, as described herein. Culture conditions for
15 expanding single antigen-specific effector cells to several billion in number with retention of antigen recognition *in vivo* are well known in the art. Such *in vitro* culture conditions typically use intermittent stimulation with antigen, often in the presence of cytokines (such as IL-2) and non-dividing feeder cells. As noted above, immunoreactive polypeptides as provided herein may be used to rapidly expand
20 antigen-specific T cell cultures in order to generate a sufficient number of cells for immunotherapy. In particular, antigen-presenting cells, such as dendritic, macrophage, monocyte, fibroblast and/or B cells, may be pulsed with immunoreactive polypeptides or transfected with one or more polynucleotides using standard techniques well known in the art. For example, antigen-presenting cells can be transfected with a
25 polynucleotide having a promoter appropriate for increasing expression in a recombinant virus or other expression system. Cultured effector cells for use in therapy must be able to grow and distribute widely, and to survive long term *in vivo*. Studies have shown that cultured effector cells can be induced to grow *in vivo* and to survive long term in substantial numbers by repeated stimulation with antigen supplemented
30 with IL-2 (*see, for example, Cheever et al., Immunological Reviews 157:177, 1997*).

Alternatively, a vector expressing a polypeptide recited herein may be introduced into antigen presenting cells taken from a patient and clonally propagated *ex vivo* for transplant back into the same patient. Transfected cells may be reintroduced into the patient using any means known in the art, preferably in sterile form by intravenous, intracavitary, intraperitoneal or intratumor administration.

Routes and frequency of administration of the therapeutic compositions described herein, as well as dosage, will vary from individual to individual, and may be readily established using standard techniques. In general, the pharmaceutical compositions and vaccines may be administered by injection (*e.g.*, intracutaneous, intramuscular, intravenous or subcutaneous), intranasally (*e.g.*, by aspiration) or orally. Preferably, between 1 and 10 doses may be administered over a 52 week period. Preferably, 6 doses are administered, at intervals of 1 month, and booster vaccinations may be given periodically thereafter. Alternate protocols may be appropriate for individual patients. A suitable dose is an amount of a compound that, when administered as described above, is capable of promoting an anti-tumor immune response, and is at least 10-50% above the basal (*i.e.*, untreated) level. Such response can be monitored by measuring the anti-tumor antibodies in a patient or by vaccine-dependent generation of cytolytic effector cells capable of killing the patient's tumor cells *in vitro*. Such vaccines should also be capable of causing an immune response that leads to an improved clinical outcome (*e.g.*, more frequent remissions, complete or partial or longer disease-free survival) in vaccinated patients as compared to non-vaccinated patients. In general, for pharmaceutical compositions and vaccines comprising one or more polypeptides, the amount of each polypeptide present in a dose ranges from about 25 μ g to 5 mg per kg of host. Suitable dose sizes will vary with the size of the patient, but will typically range from about 0.1 mL to about 5 mL.

In general, an appropriate dosage and treatment regimen provides the active compound(s) in an amount sufficient to provide therapeutic and/or prophylactic benefit. Such a response can be monitored by establishing an improved clinical outcome (*e.g.*, more frequent remissions, complete or partial, or longer disease-free survival) in treated patients as compared to non-treated patients. Increases in preexisting immune responses to a tumor protein generally correlate with an improved

clinical outcome. Such immune responses may generally be evaluated using standard proliferation, cytotoxicity or cytokine assays, which may be performed using samples obtained from a patient before and after treatment.

CANCER DETECTION AND DIAGNOSTIC COMPOSITIONS, METHODS AND KITS

5 In general, a cancer may be detected in a patient based on the presence of one or more colon tumor proteins and/or polynucleotides encoding such proteins in a biological sample (for example, blood, sera, sputum urine and/or tumor biopsies) obtained from the patient. In other words, such proteins may be used as markers to indicate the presence or absence of a cancer such as colon cancer. In addition, such
10 proteins may be useful for the detection of other cancers. The binding agents provided herein generally permit detection of the level of antigen that binds to the agent in the biological sample.

Polynucleotide primers and probes may be used to detect the level of mRNA encoding a tumor protein, which is also indicative of the presence or absence of
15 a cancer. In general, a tumor sequence should be present at a level that is at least two-fold, preferably three-fold, and more preferably five-fold or higher in tumor tissue than in normal tissue of the same type from which the tumor arose. Expression levels of a particular tumor sequence in tissue types different from that in which the tumor arose are irrelevant in certain diagnostic embodiments since the presence of tumor cells can
20 be confirmed by observation of predetermined differential expression levels, e.g., 2-fold, 5-fold, etc, in tumor tissue to expression levels in normal tissue of the same type.

Other differential expression patterns can be utilized advantageously for diagnostic purposes. For example, in one aspect of the invention, overexpression of a tumor sequence in tumor tissue and normal tissue of the same type, but not in other
25 normal tissue types, e.g. PBMCs, can be exploited diagnostically. In this case, the presence of metastatic tumor cells, for example in a sample taken from the circulation or some other tissue site different from that in which the tumor arose, can be identified and/or confirmed by detecting expression of the tumor sequence in the sample, for example using RT-PCR analysis. In many instances, it will be desired to enrich for

tumor cells in the sample of interest, e.g., PBMCs, using cell capture or other like techniques.

There are a variety of assay formats known to those of ordinary skill in the art for using a binding agent to detect polypeptide markers in a sample. *See, e.g.,* Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988. In general, the presence or absence of a cancer in a patient may be determined by (a) contacting a biological sample obtained from a patient with a binding agent; (b) detecting in the sample a level of polypeptide that binds to the binding agent; and (c) comparing the level of polypeptide with a predetermined cut-off value.

10 In a preferred embodiment, the assay involves the use of binding agent immobilized on a solid support to bind to and remove the polypeptide from the remainder of the sample. The bound polypeptide may then be detected using a detection reagent that contains a reporter group and specifically binds to the binding agent/polypeptide complex. Such detection reagents may comprise, for example, a
15 binding agent that specifically binds to the polypeptide or an antibody or other agent that specifically binds to the binding agent, such as an anti-immunoglobulin, protein G, protein A or a lectin. Alternatively, a competitive assay may be utilized, in which a polypeptide is labeled with a reporter group and allowed to bind to the immobilized binding agent after incubation of the binding agent with the sample. The extent to
20 which components of the sample inhibit the binding of the labeled polypeptide to the binding agent is indicative of the reactivity of the sample with the immobilized binding agent. Suitable polypeptides for use within such assays include full length colon tumor proteins and polypeptide portions thereof to which the binding agent binds, as described above.

25 The solid support may be any material known to those of ordinary skill in the art to which the tumor protein may be attached. For example, the solid support may be a test well in a microtiter plate or a nitrocellulose or other suitable membrane. Alternatively, the support may be a bead or disc, such as glass, fiberglass, latex or a plastic material such as polystyrene or polyvinylchloride. The support may also be a
30 magnetic particle or a fiber optic sensor, such as those disclosed, for example, in U.S. Patent No. 5,359,681. The binding agent may be immobilized on the solid support

using a variety of techniques known to those of skill in the art, which are amply described in the patent and scientific literature. In the context of the present invention, the term "immobilization" refers to both noncovalent association, such as adsorption, and covalent attachment (which may be a direct linkage between the agent and functional groups on the support or may be a linkage by way of a cross-linking agent).
5 Immobilization by adsorption to a well in a microtiter plate or to a membrane is preferred. In such cases, adsorption may be achieved by contacting the binding agent, in a suitable buffer, with the solid support for a suitable amount of time. The contact time varies with temperature, but is typically between about 1 hour and about 1 day. In
10 general, contacting a well of a plastic microtiter plate (such as polystyrene or polyvinylchloride) with an amount of binding agent ranging from about 10 ng to about 10 μ g, and preferably about 100 ng to about 1 μ g, is sufficient to immobilize an adequate amount of binding agent.

Covalent attachment of binding agent to a solid support may generally
15 be achieved by first reacting the support with a bifunctional reagent that will react with both the support and a functional group, such as a hydroxyl or amino group, on the binding agent. For example, the binding agent may be covalently attached to supports having an appropriate polymer coating using benzoquinone or by condensation of an aldehyde group on the support with an amine and an active hydrogen on the binding
20 partner (*see, e.g.,* Pierce Immunotechnology Catalog and Handbook, 1991, at A12-A13).

In certain embodiments, the assay is a two-antibody sandwich assay. This assay may be performed by first contacting an antibody that has been immobilized on a solid support, commonly the well of a microtiter plate, with the sample, such that
25 polypeptides within the sample are allowed to bind to the immobilized antibody. Unbound sample is then removed from the immobilized polypeptide-antibody complexes and a detection reagent (preferably a second antibody capable of binding to a different site on the polypeptide) containing a reporter group is added. The amount of detection reagent that remains bound to the solid support is then determined using a
30 method appropriate for the specific reporter group.

More specifically, once the antibody is immobilized on the support as described above, the remaining protein binding sites on the support are typically blocked. Any suitable blocking agent known to those of ordinary skill in the art, such as bovine serum albumin or Tween 20™ (Sigma Chemical Co., St. Louis, MO). The immobilized antibody is then incubated with the sample, and polypeptide is allowed to bind to the antibody. The sample may be diluted with a suitable diluent, such as phosphate-buffered saline (PBS) prior to incubation. In general, an appropriate contact time (*i.e.*, incubation time) is a period of time that is sufficient to detect the presence of polypeptide within a sample obtained from an individual with colon cancer at least about 95% of that achieved at equilibrium between bound and unbound polypeptide. Those of ordinary skill in the art will recognize that the time necessary to achieve equilibrium may be readily determined by assaying the level of binding that occurs over a period of time. At room temperature, an incubation time of about 30 minutes is generally sufficient.

Unbound sample may then be removed by washing the solid support with an appropriate buffer, such as PBS containing 0.1% Tween 20™. The second antibody, which contains a reporter group, may then be added to the solid support. Preferred reporter groups include those groups recited above.

The detection reagent is then incubated with the immobilized antibody-polypeptide complex for an amount of time sufficient to detect the bound polypeptide. An appropriate amount of time may generally be determined by assaying the level of binding that occurs over a period of time. Unbound detection reagent is then removed and bound detection reagent is detected using the reporter group. The method employed for detecting the reporter group depends upon the nature of the reporter group. For radioactive groups, scintillation counting or autoradiographic methods are generally appropriate. Spectroscopic methods may be used to detect dyes, luminescent groups and fluorescent groups. Biotin may be detected using avidin, coupled to a different reporter group (commonly a radioactive or fluorescent group or an enzyme). Enzyme reporter groups may generally be detected by the addition of substrate (generally for a specific period of time), followed by spectroscopic or other analysis of the reaction products.

To determine the presence or absence of a cancer, such as colon cancer, the signal detected from the reporter group that remains bound to the solid support is generally compared to a signal that corresponds to a predetermined cut-off value. In one preferred embodiment, the cut-off value for the detection of a cancer is the average mean signal obtained when the immobilized antibody is incubated with samples from patients without the cancer. In general, a sample generating a signal that is three standard deviations above the predetermined cut-off value is considered positive for the cancer. In an alternate preferred embodiment, the cut-off value is determined using a Receiver Operator Curve, according to the method of Sackett et al., *Clinical Epidemiology: A Basic Science for Clinical Medicine*, Little Brown and Co., 1985, p. 106-7. Briefly, in this embodiment, the cut-off value may be determined from a plot of pairs of true positive rates (*i.e.*, sensitivity) and false positive rates (100%-specificity) that correspond to each possible cut-off value for the diagnostic test result. The cut-off value on the plot that is the closest to the upper left-hand corner (*i.e.*, the value that encloses the largest area) is the most accurate cut-off value, and a sample generating a signal that is higher than the cut-off value determined by this method may be considered positive. Alternatively, the cut-off value may be shifted to the left along the plot, to minimize the false positive rate, or to the right, to minimize the false negative rate. In general, a sample generating a signal that is higher than the cut-off value determined by this method is considered positive for a cancer.

In a related embodiment, the assay is performed in a flow-through or strip test format, wherein the binding agent is immobilized on a membrane, such as nitrocellulose. In the flow-through test, polypeptides within the sample bind to the immobilized binding agent as the sample passes through the membrane. A second, labeled binding agent then binds to the binding agent-polypeptide complex as a solution containing the second binding agent flows through the membrane. The detection of bound second binding agent may then be performed as described above. In the strip test format, one end of the membrane to which binding agent is bound is immersed in a solution containing the sample. The sample migrates along the membrane through a region containing second binding agent and to the area of immobilized binding agent. Concentration of second binding agent at the area of immobilized antibody indicates the

presence of a cancer. Typically, the concentration of second binding agent at that site generates a pattern, such as a line, that can be read visually. The absence of such a pattern indicates a negative result. In general, the amount of binding agent immobilized on the membrane is selected to generate a visually discernible pattern when the biological sample contains a level of polypeptide that would be sufficient to generate a positive signal in the two-antibody sandwich assay, in the format discussed above. Preferred binding agents for use in such assays are antibodies and antigen-binding fragments thereof. Preferably, the amount of antibody immobilized on the membrane ranges from about 25 ng to about 1 μ g, and more preferably from about 50 ng to about 500 ng. Such tests can typically be performed with a very small amount of biological sample.

Of course, numerous other assay protocols exist that are suitable for use with the tumor proteins or binding agents of the present invention. The above descriptions are intended to be exemplary only. For example, it will be apparent to those of ordinary skill in the art that the above protocols may be readily modified to use tumor polypeptides to detect antibodies that bind to such polypeptides in a biological sample. The detection of such tumor protein specific antibodies may correlate with the presence of a cancer.

A cancer may also, or alternatively, be detected based on the presence of T cells that specifically react with a tumor protein in a biological sample. Within certain methods, a biological sample comprising CD4⁺ and/or CD8⁺ T cells isolated from a patient is incubated with a tumor polypeptide, a polynucleotide encoding such a polypeptide and/or an APC that expresses at least an immunogenic portion of such a polypeptide, and the presence or absence of specific activation of the T cells is detected. Suitable biological samples include, but are not limited to, isolated T cells. For example, T cells may be isolated from a patient by routine techniques (such as by Ficoll/Hypaque density gradient centrifugation of peripheral blood lymphocytes). T cells may be incubated *in vitro* for 2-9 days (typically 4 days) at 37°C with polypeptide (e.g., 5 - 25 μ g/ml). It may be desirable to incubate another aliquot of a T cell sample in the absence of tumor polypeptide to serve as a control. For CD4⁺ T cells, activation is preferably detected by evaluating proliferation of the T cells. For CD8⁺ T cells,

activation is preferably detected by evaluating cytolytic activity. A level of proliferation that is at least two fold greater and/or a level of cytolytic activity that is at least 20% greater than in disease-free patients indicates the presence of a cancer in the patient.

5 As noted above, a cancer may also, or alternatively, be detected based on the level of mRNA encoding a tumor protein in a biological sample. For example, at least two oligonucleotide primers may be employed in a polymerase chain reaction (PCR) based assay to amplify a portion of a tumor cDNA derived from a biological sample, wherein at least one of the oligonucleotide primers is specific for (*i.e.*,
10 hybridizes to) a polynucleotide encoding the tumor protein. The amplified cDNA is then separated and detected using techniques well known in the art, such as gel electrophoresis.

Similarly, oligonucleotide probes that specifically hybridize to a polynucleotide encoding a tumor protein may be used in a hybridization assay to detect
15 the presence of polynucleotide encoding the tumor protein in a biological sample.

To permit hybridization under assay conditions, oligonucleotide primers and probes should comprise an oligonucleotide sequence that has at least about 60%, preferably at least about 75% and more preferably at least about 90%, identity to a portion of a polynucleotide encoding a tumor protein of the invention that is at least 10
20 nucleotides, and preferably at least 20 nucleotides, in length. Preferably, oligonucleotide primers and/or probes hybridize to a polynucleotide encoding a polypeptide described herein under moderately stringent conditions, as defined above. Oligonucleotide primers and/or probes which may be usefully employed in the diagnostic methods described herein preferably are at least 10-40 nucleotides in length.
25 In a preferred embodiment, the oligonucleotide primers comprise at least 10 contiguous nucleotides, more preferably at least 15 contiguous nucleotides, of a DNA molecule having a sequence as disclosed herein. Techniques for both PCR based assays and hybridization assays are well known in the art (*see*, for example, Mullis et al., *Cold Spring Harbor Symp. Quant. Biol.*, 51:263, 1987; Erlich ed., *PCR Technology*, Stockton
30 Press, NY, 1989).

One preferred assay employs RT-PCR, in which PCR is applied in conjunction with reverse transcription. Typically, RNA is extracted from a biological sample, such as biopsy tissue, and is reverse transcribed to produce cDNA molecules. PCR amplification using at least one specific primer generates a cDNA molecule, which may be separated and visualized using, for example, gel electrophoresis. Amplification may be performed on biological samples taken from a test patient and from an individual who is not afflicted with a cancer. The amplification reaction may be performed on several dilutions of cDNA spanning two orders of magnitude. A two-fold or greater increase in expression in several dilutions of the test patient sample as compared to the same dilutions of the non-cancerous sample is typically considered positive.

In another aspect of the present invention, cell capture technologies may be used in conjunction, with, for example, real-time PCR to provide a more sensitive tool for detection of metastatic cells expressing colon tumor antigens. Detection of colon cancer cells in biological samples, e.g., bone marrow samples, peripheral blood, and small needle aspiration samples is desirable for diagnosis and prognosis in colon cancer patients.

Immunomagnetic beads coated with specific monoclonal antibodies to surface cell markers, or tetrameric antibody complexes, may be used to first enrich or positively select cancer cells in a sample. Various commercially available kits may be used, including Dynabeads® Epithelial Enrich (DynaL Biotech, Oslo, Norway), StemSep™ (StemCell Technologies, Inc., Vancouver, BC), and RosetteSep (StemCell Technologies). A skilled artisan will recognize that other methodologies and kits may also be used to enrich or positively select desired cell populations. Dynabeads® Epithelial Enrich contains magnetic beads coated with mAbs specific for two glycoprotein membrane antigens expressed on normal and neoplastic epithelial tissues. The coated beads may be added to a sample and the sample then applied to a magnet, thereby capturing the cells bound to the beads. The unwanted cells are washed away and the magnetically isolated cells eluted from the beads and used in further analyses.

RosetteSep can be used to enrich cells directly from a blood sample and consists of a cocktail of tetrameric antibodies that targets a variety of unwanted cells

and crosslinks them to glycophorin A on red blood cells (RBC) present in the sample, forming rosettes. When centrifuged over Ficoll, targeted cells pellet along with the free RBC. The combination of antibodies in the depletion cocktail determines which cells will be removed and consequently which cells will be recovered. Antibodies that are available include, but are not limited to: CD2, CD3, CD4, CD5, CD8, CD10, CD11b, CD14, CD15, CD16, CD19, CD20, CD24, CD25, CD29, CD33, CD34, CD36, CD38, CD41, CD45, CD45RA, CD45RO, CD56, CD66B, CD66e, HLA-DR, IgE, and TCR $\alpha\beta$.

Additionally, it is contemplated in the present invention that mAbs specific for colon tumor antigens can be generated and used in a similar manner. For example, mAbs that bind to tumor-specific cell surface antigens may be conjugated to magnetic beads, or formulated in a tetrameric antibody complex, and used to enrich or positively select metastatic colon tumor cells from a sample. Once a sample is enriched or positively selected, cells may be lysed and RNA isolated. RNA may then be subjected to RT-PCR analysis using colon tumor-specific primers in a real-time PCR assay as described herein. One skilled in the art will recognize that enriched or selected populations of cells may be analyzed by other methods (*e.g. in situ* hybridization or flow cytometry).

In another embodiment, the compositions described herein may be used as markers for the progression of cancer. In this embodiment, assays as described above for the diagnosis of a cancer may be performed over time, and the change in the level of reactive polypeptide(s) or polynucleotide(s) evaluated. For example, the assays may be performed every 24-72 hours for a period of 6 months to 1 year, and thereafter performed as needed. In general, a cancer is progressing in those patients in whom the level of polypeptide or polynucleotide detected increases over time. In contrast, the cancer is not progressing when the level of reactive polypeptide or polynucleotide either remains constant or decreases with time.

Certain *in vivo* diagnostic assays may be performed directly on a tumor. One such assay involves contacting tumor cells with a binding agent. The bound binding agent may then be detected directly or indirectly via a reporter group. Such

binding agents may also be used in histological applications. Alternatively, polynucleotide probes may be used within such applications.

As noted above, to improve sensitivity, multiple tumor protein markers may be assayed within a given sample. It will be apparent that binding agents specific
5 for different proteins provided herein may be combined within a single assay. Further, multiple primers or probes may be used concurrently. The selection of tumor protein markers may be based on routine experiments to determine combinations that results in optimal sensitivity. In addition, or alternatively, assays for tumor proteins provided herein may be combined with assays for other known tumor antigens.

10 The present invention further provides kits for use within any of the above diagnostic methods. Such kits typically comprise two or more components necessary for performing a diagnostic assay. Components may be compounds, reagents, containers and/or equipment. For example, one container within a kit may contain a monoclonal antibody or fragment thereof that specifically binds to a tumor
15 protein. Such antibodies or fragments may be provided attached to a support material, as described above. One or more additional containers may enclose elements, such as reagents or buffers, to be used in the assay. Such kits may also, or alternatively, contain a detection reagent as described above that contains a reporter group suitable for direct or indirect detection of antibody binding.

20 Alternatively, a kit may be designed to detect the level of mRNA encoding a tumor protein in a biological sample. Such kits generally comprise at least one oligonucleotide probe or primer, as described above, that hybridizes to a polynucleotide encoding a tumor protein. Such an oligonucleotide may be used, for example, within a PCR or hybridization assay. Additional components that may be
25 present within such kits include a second oligonucleotide and/or a diagnostic reagent or container to facilitate the detection of a polynucleotide encoding a tumor protein.

The following Examples are offered by way of illustration and not by way of limitation.

EXAMPLES

EXAMPLE 1

IDENTIFICATION OF COLON TUMOR PROTEIN cDNAs

5 This Example illustrates the identification of cDNA molecules encoding colon tumor proteins using PCR-based cDNA subtraction methodology.

 A modification of the Clontech (Palo Alto, CA) PCR-SelectTM cDNA subtraction methodology was employed to obtain cDNA populations enriched in cDNAs derived from transcripts that are differentially expressed in colon tumor
10 samples. By this methodology, mRNA populations were isolated from colon tumor and metastatic tumor samples ("tester" mRNA) as well as from normal tissues, such as brain, pancreas, bone marrow, liver, heart, lung, stomach and small intestine ("driver" mRNA). From the tester and driver mRNA populations, cDNA was synthesized by standard methodology. See, e.g., Ausubel, F.M. et al., *Short Protocols in Molecular*
15 *Biology* (4th ed., John Wiley and Sons, Inc., 1999).

 The subtraction steps were performed using a PCR-based protocol that was modified to generate fragments larger than would be derived by the Clontech methodology. By this modified protocol, the tester and driver cDNAs were separately digested with five restriction endonucleases (Mlu I, Msc I, Pvu II, Sal I and Stu I) each
20 of which recognize a unique 6-base pair nucleotide sequence. This digestion resulted in an average cDNA size of 600 bp, rather than the average size of 300 bp that results from digestion with Rsa I according to the Clontech methodology. This modification did not affect the ultimate subtraction efficiency.

 Following the restriction digestion, adapter oligonucleotides having
25 unique nucleotide sequences were ligated onto the 5' ends of the tester cDNAs; adapter oligonucleotides were not ligated onto the driver cDNAs. The tester and driver cDNAs were subsequently hybridized one to the other using an excess of driver cDNA. This hybridization step resulted in populations of (a) unhybridized tester cDNAs, (b) tester cDNAs hybridized to other tester cDNAs, (c) tester cDNAs hybridized to driver
30 cDNAs, (d) unhybridized driver cDNAs and (e) driver cDNAs hybridized to driver cDNAs.

Tester cDNAs hybridized to other tester cDNAs were selectively amplified by a polymerase chain reaction (PCR) employing primers complementary to the ligated adapters. Because only tester cDNAs were ligated to adapter sequences, neither unhybridized tester or driver cDNAs, tester cDNAs hybridized to driver cDNAs nor driver cDNAs hybridized to driver cDNAs were amplified using adapter specific oligonucleotides. The PCR amplified tester cDNAs were cloned into the pCR2.1 plasmid vector (Invitrogen; Carlsbad, CA) to create a libraries enriched in differentially expressed colon tumor antigen and colon metastatic tumor antigen specific cDNAs.

Three thousand clones from the pCR2.1 tumor antigen cDNA libraries were randomly selected and used to obtain clones for microarray analysis (performed by Rosetta; Seattle, WA) and nucleotide sequencing. The cDNA insert from each pCR2.1 clone was PCR amplified as follows. Briefly, 0.5 μ l of glycerol stock solution was added to 99.5 μ l of PCR mix containing 80 μ l H₂O, 10 μ l 10X PCR Buffer, 6 μ l MgCl₂, 1 μ l 10 mM dNTPs, 1 μ l 100 mM M13 forward primer (CACGACGTTGTAAAACGACGG), 1 μ l 100 mM M13 reverse primer (CACAGGAAACAGCTATGACC), and 0.5 μ l 5 u/ml Taq DNA polymerase. The M13 forward and reverse primers used herein were obtained from Operon Technologies (Alameda, CA). The PCR amplification was performed for thirty cycles under the following conditions: 95°C for 5 minutes, 92°C for 30 seconds, 57°C for 40 seconds, 75°C for 2 minutes and 75°C for 5 minutes.

mRNA expression levels for representative clones were determined using microarray technology in colon tumor tissues (n=25), normal colon tissues (n=6), kidney, lung, liver, brain, heart, esophagus, small intestine, stomach, pancreas, adrenal gland, salivary gland, resting PBMC, activated PBMC, bone marrow, dendritic cells, spinal cord, blood vessels, skeletal muscle, skin, breast and fetal tissues. An exemplary methodology for performing the microarray analysis is described in Schena *et al.*, *Science* 270:467-470. The number of tissue samples tested in each case was one (n=1), except where specifically noted above; additionally, all the above-mentioned tissues were derived from humans.

The PCR amplification products were dotted onto slides in an array format, with each product occupying a unique location in the array. mRNA was

extracted from the tissue sample to be tested, and fluorescent-labeled cDNA probes were generated by reverse transcription, according to standard methodology, in the presence of fluorescent nucleotides $\psi 5$ and $\psi 3$. See, e.g., Ausubel, et al., *supra* for exemplary reaction conditions for performing the reverse transcription reaction; $\psi 5$ and $\psi 3$ fluorescent labeled nucleotides may be obtained, e.g., from Amersham Pharmacia (Uppsala, Sweden) or NEN® Life Science Products, Inc. (Boston, MA). The microarrays were probed with the fluorescent-labeled cDNAs, the slides were scanned and fluorescence intensity was measured. Genetic MicroSystems instrumentation for preparing the cDNA microarrays and for measuring fluorescence intensity is available from Affymetrix (Santa Clara, CA).

An elevated fluorescence intensity in a microarray sector probed with cDNA probes obtained from a colon tumor or colon metastatic tumor tissue as compared to the fluorescence intensity in the same sector probed with cDNA probes obtained from a normal tissue indicates a tumor antigen gene that is differentially expressed in colon tumor or colon metastatic tumor tissue.

Clones disclosed herein as SEQ ID NOs: 1-234 and described in Tables 2-4 were identified from the PCR subtracted differential colon tumor and colon metastatic tumor cDNA libraries by the microarray based methodology. Of these 234 clones, those corresponding to SEQ ID NOs: 1, 6, 18-20, 27, 30, 37, 40, 57, 65, 81, 82, 86, 88, 91, 95, 96, 106, 107, 117, 121, 123, 126, 130, 148, 150, 152, 155, 157, 159, 161, 174, 175, 180, 182, 187, 190, 191, 192, 203, 204 and 209 showed no significant similarity to known sequences in Genbank.

TABLE 2

CDNA SEQUENCES SHOWING NO SIGNIFICANT SIMILARITY TO SEQUENCE IN GENBANK

| Clone | SEQ ID NO. | EST | Element (384) | Element (96) | Ratio | Median Signal 1 | Median Signal 2 | 96 Well Location |
|-------|------------|----------------------------|---------------|--------------|-------|-----------------|-----------------|------------------|
| 54172 | 1 | Parathyroid/breast | p0022r16c12 | R0085 H6 | 3.24 | 0.276 | 0.085 | 5G12 |
| 54034 | 6 | Ovarian | p0018r08c10 | R0067 H5 | 2.24 | 0.179 | 0.08 | 4D6 |
| 53949 | 18 | Colon/pancreatic islet | p0016r15c12 | R0061 F6 | 2.32 | 0.145 | 0.062 | 3E 5 |
| 53898 | 19 | Colon/Gastric | p0016r01c14 | R0058 B7 | 4.43 | 0.423 | 0.095 | 3A2 |
| 54069 | 20 | Prostate/colon | p0019r03c02 | R0070 F1 | 2.5 | 0.136 | 0.054 | 4G5 |
| 54089 | 27 | Colon/HCC cell line | p0019r14c18 | R0073 D9 | 2.97 | 0.096 | 0.032 | 5A1 |
| 54181 | 30 | Br/Li/Ut/Pr | p0023r09c19 | R0088 A10 | 2.85 | 0.264 | 0.092 | 5H9 |
| 54147 | 37 | Colon only | p0021r12c01 | R0080 G1 | 2.05 | 0.132 | 0.064 | 5E 11 |
| 54039 | 40 | Ovary | p0018r09c06 | R0068 B3 | 2.03 | 0.185 | 0.091 | 4D11 |
| 54059 | 57 | Novel | p0018r13c20 | R0069 B10 | 2.02 | 0.089 | 0.044 | 4F7 |
| 54141 | 65 | HCC cell line/colon/testis | p0021r07c03 | R0079 E2 | 2.35 | 0.106 | 0.045 | 5E 5 |
| 54120 | 81 | Novel | p0020r11c07 | R0076 E4 | 2.02 | 0.087 | 0.043 | 5C8 |
| 54145 | 82 | Ut/Plac/Br/Pr | p0021r11c01 | R0080 E1 | 2.5 | 0.147 | 0.059 | 5E 9 |
| 54152 | 86 | Ut/Lu/Co/Pancreatic islet | p0021r14c23 | R0081 C12 | 2.14 | 0.141 | 0.066 | 5F4 |
| 54146 | 88 | Br/Co/melanocyte | p0021r11c19 | R0080 E10 | 2.07 | 0.097 | 0.047 | 5E 10 |
| 54020 | 91 | Fetal liver/heart | p0017r16c12 | R0065 H6 | 2.18 | 0.133 | 0.061 | 4C4 |
| 54161 | 95 | Fetal liver spleen | p0022r05c16 | R0083 B8 | 2.07 | 0.083 | 0.04 | 5G1 |
| 54162 | 96 | Lot EST | p0022r05c22 | R0083 B11 | 3.74 | 0.205 | 0.055 | 5G2 |
| 54098 | 106 | Lot EST | p0020r02c05 | R0074 C3 | 2.06 | 0.064 | 0.031 | 5A10 |
| 54173 | 107 | Co/Pan/Kid/Liver | p0022r16c23 | R0085 G12 | 2.62 | 0.14 | 0.053 | 5H1 |

| Clone | SEQ ID NO. | EST | Element (384) | Element (96) | Ratio | Median Signal 1 | Median Signal 2 | 96 Well Location |
|-------|------------|-----------------------------|---------------|--------------|-------|-----------------|-----------------|------------------|
| 54183 | 117 | Co/Bm/Ut/Lu | p0023r10c20 | R0088 D10 | 2.8 | 0.092 | 0.033 | 5H11 |
| 53918 | 121 | Infant brain/breast | p0016r07c15 | R0059 E8 | 2.06 | 0.104 | 0.051 | 3B10 |
| 53910 | 123 | Co/Ut | p0016r05c11 | R0059 A6 | 2.01 | 0.098 | 0.049 | 3B2 |
| 53917 | 126 | Infant brain/gall bladder | p0016r07c02 | R0059 F1 | 2 | 0.102 | 0.051 | 3B9 |
| 53999 | 130 | Kid/Thymus/Co | p0017r12c08 | R0064 H4 | 2.75 | 0.269 | 0.098 | 4A7 |
| 54074 | 148 | Pr | p0019r04c04 | R0070 H2 | 2 | 0.198 | 0.099 | 4G10 |
| 53961 | 150 | Novel | p0017r03c06 | R0062 F3 | 3.45 | 0.069 | 0.02 | 3F5 |
| 53933 | 152 | Lot EST | p0016r10c21 | R0060 C11 | 2.64 | 0.14 | 0.053 | 3D1 |
| 53924 | 155 | Novel | p0016r08c11 | R0059 G6 | 3.14 | 0.144 | 0.046 | 3C4 |
| 54068 | 157 | Lot EST | p0019r01c12 | R0070 B6 | 2.01 | 0.182 | 0.091 | 4G4 |
| 53959 | 159 | Germinal center B cell | p0017r03c01 | R0062 E1 | 2.01 | 0.042 | 0.021 | 3F3 |
| 53931 | 161 | Pr/Lu | p0016r10c17 | R0060 C9 | 2.41 | 0.152 | 0.063 | 3C11 |
| 54091 | 174 | Kid/Stomach | p0019r15c06 | R0073 F3 | 2.1 | 0.076 | 0.036 | 5A3 |
| 54013 | 175 | Fetal tissues/testis | p0017r15c03 | R0065 E2 | 2.32 | 0.183 | 0.079 | 4B9 |
| 53963 | 180 | Lot EST | p0017r03c12 | R0062 F6 | 2.59 | 0.256 | 0.099 | 3F7 |
| 54067 | 182 | Lot EST | p0018r16c20 | R0069 H10 | 4.8 | 0.347 | 0.072 | 4G3 |
| 53966 | 187 | Infant brain | p0017r04c07 | R0062 G4 | 2.08 | 0.119 | 0.057 | 3F10 |
| 54094 | 190 | Co/Fetal retina | p0019r16c01 | R0073 G1 | 2.11 | 0.149 | 0.071 | 5A6 |
| 53977 | 191 | 1887043 | p0017r05c12 | R0063 B6 | 2.35 | 0.164 | 0.07 | 3G9 |
| 54123 | 192 | Infant brain/multiple scler | p0020r15c04 | R0077 F2 | 2.01 | 0.091 | 0.045 | 5C11 |
| 54016 | 203 | Novel | p0017r15c16 | R0065 F8 | 2.04 | 0.113 | 0.055 | 4B12 |
| 54018 | 204 | Br/Co | p0017r15c23 | R0065 E12 | 3.48 | 0.203 | 0.058 | 4C2 |
| 53988 | 209 | Kid/Co/Fetal brain | p0017r08c20 | R0063 H10 | 2.88 | 0.117 | 0.041 | 3H8 |

TABLE 3

SEQUENCES WITH SOME DEGREE OF SIMILARITY TO SEQUENCES IN GENBANK WITH NO KNOWN FUNCTION

| Clone | SEQ ID NO. | Genbank | EST | Element (384) | Element (96) | Ratio | Median Signal 1 | Median Signal 2 | 96 Well Location |
|-------|------------|---------------------------------------|------------------------|---------------|--------------|-------|-----------------|-----------------|------------------|
| 54104 | 2 | PAC 75N13 on chromosome Xq21.1 | Colon only | p0020r03c18 | R0074 F9 | 2.15 | 0.098 | 0.045 | 5B4 |
| 54149 | 5 | cDNA FLJ10461 fis, clone NT2RP1001482 | Ovarian | p0021r13c12 | R0081 B6 | 2.5 | 0.068 | 0.027 | 5F1 |
| 53948 | 8 | 12p12 BAC RPC111-267J23 | Testis/colon/liver | p0016r15c11 | R0061 E6 | 2.05 | 0.147 | 0.072 | 3E 4 |
| 54026 | 9 | Clone 164F3 on chromosome Xq21.33-23 | Fetal liver/lung/colon | p0018r04c10 | R0066 H5 | 2 | 0.125 | 0.062 | 4C10 |
| 54174 | 17 | PAC clone RP1-170019 from 7p15-p21 | Colon only | p0023r03c09 | R0086 E5 | 2.63 | 0.221 | 0.084 | 5H2 |

| Clone | SEQ ID NO. | Genbank | EST | Element (384) | Element (96) | Ratio | Median Signal 1 | Median Signal 2 | 96 Well Location |
|-------|------------|---|------------------------------|---------------|--------------|-------|-----------------|-----------------|------------------|
| 54048 | 21 | cDNA FLJ20676 fis, clone KAIA4294 | Pancreatic islet/prostate | p0018r11c17 | R0068 E9 | 5.15 | 0.315 | 0.061 | 4E 8 |
| 54031 | 22 | Chromosome 17, clone hRPC.1171_I 10 | Co/Pr/Ov/Ut | p0018r07c23 | R0067 E12 | 4.66 | 0.454 | 0.098 | 4D3 |
| 54079 | 31 | PAC 75N13 on chromosome Xq21.1 | Co/Gas | p0019r06c18 | R0071 D9 | 3.04 | 0.199 | 0.066 | 4H3 |
| 54160 | 33 | Clone 146H21 on chromosome Xq22 | Colon only | p0022r05c08 | R0083 B4 | 3.7 | 0.215 | 0.058 | 5F12 |
| 54078 | 35 | PAC 75N13 on chromosome Xq21.1 | Colon only | p0019r06c09 | R0071 C5 | 2.79 | 0.145 | 0.052 | 4H2 |
| 54037 | 41 | Constitutive fragile region FRA3B sequence 90% | Pancreatic islet/colon | p0018r08c24 | R0067 H12 | 2.37 | 0.128 | 0.054 | 4D9 |

| Clone | SEQ ID NO. | Genbank | EST | Element (384) | Element (96) | Ratio | Median Signal 1 | Median Signal 2 | 96 Well Location |
|-------|------------|---|---------------------|---------------|--------------|-------|-----------------|-----------------|------------------|
| 54052 | 51 | cDNA FLJ10610 fis, clone NT2RP20052 93 | Novel | p0018r12c21 | R0068 G11 | 2.36 | 0.072 | 0.031 | 4E 12 |
| 54124 | 63 | Clone RP1- 39G22 on chromosome 1p32.1-34.3 | Kid/Ut/Infant brain | p0020r16c10 | R0077 H5 | 2.07 | 0.149 | 0.072 | 5C12 |
| 54065 | 69 | cDNA FLJ10969 fis, clone PLACE10009 09 | Kid/Ut | p0018r15c19 | R0069 E10 | 2.36 | 0.193 | 0.082 | 4G1 |
| 54060 | 70 | BAC clone 215012 | Pancreatic islet | p0018r14c16 | R0069 D8 | 2.15 | 0.099 | 0.046 | 4F8 |
| 54136 | 78 | KIAA1077 protein | Bt/Pr/Ut | p0021r04c24 | R0078 H12 | 2.27 | 0.112 | 0.049 | 5D12 |
| 54140 | 80 | PAC 454G6 on chromosome 1q24 | Pan/HeLa cell/Ut | p0021r06c08 | R0079 D4 | 2.17 | 0.062 | 0.029 | 5E 4 |
| 54117 | 83 | KIAA0152 | Ut/Co/Br/Lu | p0020r10c13 | R0076 C7 | 2.02 | 0.063 | 0.031 | 5C5 |
| 54159 | 90 | cDNA DKFZp586O 0118 | Lot | p0022r04c08 | R0082 H4 | 2.64 | 0.159 | 0.06 | 5F11 |

| Clone | SEQ ID NO. | Genbank | EST | Element (384) | Element (96) | Ratio | Median Signal 1 | Median Signal 2 | 96 Well Location |
|-------|------------|---------------------------------------|--------------------------|---------------|--------------|-------|-----------------|-----------------|------------------|
| 54030 | 94 | CGI-151/KIAA0992 protein | Endothelial cell/Sk Musc | p0018r06c22 | R0067 D11 | 2.02 | 0.154 | 0.076 | 4D2 |
| 54133 | 101 | cDNA DKFZp586I2022 | Lu/Co/Ut | p0021r04c02 | R0078 H1 | 2.63 | 0.136 | 0.052 | 5D9 |
| 54131 | 104 | cDNA FLJ10549 fis, clone NT2RP2001976 | Uv/GC/Pr | p0021r03c08 | R0078 F4 | 2.03 | 0.083 | 0.041 | 5D7 |
| 54122 | 105 | cDNA DKFZp434C0523 | Embryo/fetal brain | p0020r12c04 | R0076 H2 | 2.36 | 0.224 | 0.095 | 5C10 |
| 54179 | 110 | cDNA FLJ10610 fis, clone NT2RP2005293 | Thymus/fetal heart | p0023r08c18 | R0087 H9 | 2.13 | 0.089 | 0.042 | 5H7 |
| 54027 | 116 | cDNA FLJ10884 fis, clone NT2RP4001950 | GC/testis | p0018r05c06 | R0067 B3 | 2.15 | 0.181 | 0.084 | 4C11 |
| 54106 | 119 | KIAA1289 | Fetal tissue/melanocyte | p0020r04c19 | R0074 G10 | 2.09 | 0.104 | 0.05 | 5B6 |

| Clone | SEQ ID NO. | Genbank | EST | Element (384) | Element (96) | Ratio | Median Signal 1 | Median Signal 2 | 96 Well Location |
|-------|------------|--|-------------------------|---------------|--------------|-------|-----------------|-----------------|------------------|
| 53904 | 122 | Chromosome 17, clone hRPK.692_E_18 | Co/fetal/placenta | p0016r03c15 | R0058 E8 | 4.59 | 0.445 | 0.097 | 3A8 |
| 53903 | 124 | cDNA FLJ10823 fis, clone NT2RP40010_80 | Colon only | p0016r03c12 | R0058 F6 | 2.08 | 0.111 | 0.053 | 3A7 |
| 53928 | 133 | citb_338_f_2 4, complete sequence | Ut/infant brain | p0016r09c19 | R0060 A10 | 3.14 | 0.166 | 0.053 | 3C8 |
| 53930 | 139 | Chromosome 19 | 6882084/6893421 | p0016r10c04 | R0060 D2 | 2.35 | 0.127 | 0.054 | 3C10 |
| 54005 | 143 | Chromosome 5 clone CTC-436P18 | GCB/infant brain | p0017r12c22 | R0064 H11 | 2.07 | 0.132 | 0.064 | 4B1 |
| 54083 | 146 | 12q24 PAC RPCII-261P5 | Novel | p0019r08c18 | R0071 H9 | 2.12 | 0.057 | 0.027 | 4H7 |
| 54105 | 149 | Clone RP4-621F18 on chromosome 1p11.4-21.3 | Total fetus/fetal liver | p0020r04c18 | R0074 H9 | 2.46 | 0.095 | 0.039 | 5B5 |
| 53906 | 154 | cDNA FLJ10679 fis, clone NT2RP20065_65 | Lot EST | p0016r03c24 | R0058 F12 | 2.04 | 0.13 | 0.064 | 3A10 |

| Clone | SEQ ID NO. | Genbank | EST | Element (384) | Element (96) | Ratio | Median Signal 1 | Median Signal 2 | 96 Well Location |
|-------|------------|--|---------------------------|---------------|--------------|-------|-----------------|-----------------|------------------|
| 53942 | 160 | KIAA1050 | Fetus/fetal lung | p0016r14c05 | R0061 C3 | 2.02 | 0.067 | 0.033 | 3D10 |
| 53935 | 162 | cDNA FLJ11127 fis, clone PLACE10062 25 | Co/Pan/Ov/Ut | p0016r11c08 | R0060 F4 | 2.77 | 0.19 | 0.069 | 3D3 |
| 54000 | 165 | KIAA0965 | Fetus/Co/Ut | p0017r12c09 | R0064 G5 | 2.12 | 0.149 | 0.07 | 4A8 |
| 53953 | 169 | cDNA DKFZp586H 0519 | Ovary/fetal brain | p0016r15c24 | R0061 F12 | 2.49 | 0.141 | 0.057 | 3E9 |
| 53945 | 173 | cDNA FLJ20127 fis, clone COL06176 | Novel | p0016r14c20 | R0061 D10 | 2.21 | 0.108 | 0.049 | 3E1 |
| 53987 | 178 | Clone RP1- 155G6 on chromosome 20 | HeLa/placenta/testis | p0017r08c16 | R0063 H8 | 2.05 | 0.159 | 0.078 | 3H7 |
| 54057 | 183 | PAC RPCI-1 133G21 map 21q11.1 region D21S190 | Novel | p0018r13c11 | R0069 A6 | 2.11 | 0.091 | 0.043 | 4F5 |
| 53960 | 193 | BAC clone RG083M05 from 7q21- 7q22 | Subtracted Hippocampus | p0017r03c02 | R0062 F1 | 2.48 | 0.07 | 0.028 | 3F4 |

| Clone | SEQ ID NO. | Genbank | EST | Element (384) | Element (96) | Ratio | Median Signal 1 | Median Signal 2 | 96 Well Location |
|-------|------------|--|-----------------|---------------|--------------|-------|-----------------|-----------------|------------------|
| 53976 | 194 | Human STS WI-14644 | | p0017r05c09 | R0063 A5 | 2.53 | 0.243 | 0.096 | 3G8 |
| 54081 | 199 | PAC RPCI-1 133G21 map 21q11.1 region D21S190 | Colon only | p0019r07c10 | R0071 F5 | 4.66 | 0.225 | 0.048 | 4H5 |
| 54082 | 200 | PAC clone RP5-118517 from 7q11.23-q21 | GCB/total fetus | p0019r07c16 | R0071 F8 | 2.38 | 0.105 | 0.044 | 4H6 |
| 53992 | 202 | cDNA FLJ20673 fs, clone KAIA4464 | Kid/GCB/Co | p0017r11c08 | R0064 F4 | 2.03 | 0.128 | 0.063 | 3H12 |
| 53973 | 206 | KIAA0715 | Colon/Brain | p0017r04c24 | R0062 H12 | 4.39 | 0.196 | 0.045 | 3G5 |
| 53982 | 208 | KIAA1225 | Lym/Co | p0017r06c24 | R0063 D12 | 2.22 | 0.107 | 0.048 | 3H2 |
| 53991 | 211 | cDNA FLJ20171 fs, clone COL09761 | Lu/Ut/Ct | p0017r10c21 | R0064 C11 | 2.81 | 0.062 | 0.022 | 3H11 |

TABLE 4

CDNA SEQUENCES WITH SOME DEGREE OF SIMILARITY TO KNOWN SEQUENCES IN GENBANK

| Clone | SEQ ID NO. | Genbank | EST | Element (384) | Element (96) | Ratio | Median Signal 1 | Median Signal 2 | 96 Well Location |
|-------|------------|---|-----|---------------|--------------|-------|-----------------|-----------------|------------------|
| 53978 | 3 | Glutamine:fructose-6-phosphate amidotransferase | | p0017r05c14 | R0063 B7 | 3.24 | 0.182 | 0.056 | 3G10 |
| 54184 | 4 | Colon Kruppel-like factor | | p0023r10c22 | R0088 D11 | 3.55 | 0.222 | 0.062 | 5H12 |
| 54085 | 7 | Human beta 2 gene | | p0019r11c24 | R0072 F12 | 2.08 | 0.184 | 0.089 | 4H9 |
| 53907 | 10 | Lysyl hydroxylase isoform 2 | | p0016r04c04 | R0058 H2 | 2.25 | 0.218 | 0.097 | 3A11 |
| 54066 | 11 | Mucin 11 | | p0018r15c23 | R0069 E12 | 3.87 | 0.222 | 0.057 | 4G2 |
| 54017 | 12 | Mucin 11 | | p0017r15c20 | R0065 F10 | 5.21 | 0.241 | 0.046 | 4C1 |
| 54006 | 13 | Mucin 11 | | p0017r13c10 | R0065 B5 | 3.97 | 0.246 | 0.062 | 4B2 |
| 53962 | 14 | Epiregulin (EGF family) | | p0017r03c09 | R0062 E5 | 2.61 | 0.083 | 0.032 | 3F6 |
| 54028 | 15 | Mucin 12 | | p0018r05c15 | R0067 A8 | 2.14 | 0.068 | 0.032 | 4C12 |
| 54166 | 16 | E1A enhancer binding protein | | p0022r10c04 | R0084 D2 | 2.5 | 0.226 | 0.09 | 5G6 |

| Clone | SEQ ID NO. | Genbank | EST | Element (384) | Element (96) | Ratio | Median Signal 1 | Median Signal 2 | 96 Well Location |
|-------|------------|--|----------------|---------------|--------------|-------|-----------------|-----------------|------------------|
| 54154 | 23 | Alpha topoisomerase truncated form | | p0021r15c12 | R0081 F6 | 3.22 | 0.315 | 0.098 | 5F6 |
| 54009 | 24 | Cytokeratin 20 | | p0017r14c11 | -R0065 C6 | 4.07 | 0.185 | 0.045 | 4B5 |
| 54070 | 25 | Erythroblastosis virus oncogene homolog 2 | | p0019r03c03 | R0070 E2 | 2.05 | 0.172 | 0.084 | 4G6 |
| 53998 | 26 | Polyadenylate binding protein II | | p0017r12c07 | R0064 G4 | 3.73 | 0.368 | 0.099 | 4A6 |
| 54182 | 28 | Transforming growth factor-beta induced gene product | | p0023r10c07 | R0088 C4 | 3.14 | 0.21 | 0.067 | 5H10 |
| 53989 | 29 | GDP-mannose 4,6 dehydratase | | p0017r08c24 | R0063 H12 | 3.77 | 0.259 | 0.069 | 3H9 |
| 54114 | 32 | Mus fork head transcription factor gene 92% | Kid/Co/Lu/UvPr | p0020r09c13 | R0076 A7 | 3.39 | 0.185 | 0.055 | 5C2 |

| Clone | SEQ ID NO. | Genbank | EST | Element (384) | Element (96) | Ratio | Median Signal 1 | Median Signal 2 | 96 Well Location |
|-------|------------|---|-----|---------------|--------------|-------|-----------------|-----------------|------------------|
| 54168 | 34 | Glutamine:fructose-6-phosphate amidotransferase | | p0022r15c16 | R0085 F8 | 2.4 | 0.224 | 0.093 | 5G8 |
| 53900 | 36 | Intestinal peptide-associated transporter HPT-1 | | p0016r03c01 | R0058 E1 | 2.11 | 0.114 | 0.054 | 3A4 |
| 54033 | 38 | Human proteinase activated receptor-2 | | p0018r08c07 | R0067 G4 | 2.89 | 0.143 | 0.049 | 4D5 |
| 54022 | 39 | GalNAc-T3 gene | | p0017r16c21 | R0065 G11 | 2.54 | 0.193 | 0.076 | 4C6 |
| 54129 | 42 | CD24 signal transducer gene | | p0021r02c15 | R0078 C8 | 2.5 | 0.239 | 0.096 | 5D5 |
| 54054 | 43 | Human c-myc gene | | p0018r13c02 | R0069 B1 | 3.15 | 0.282 | 0.089 | 4F2 |
| 54055 | 44 | Pyroline-5-carboxylate synthase long form | | p0018r13c03 | R0069 A2 | 2.01 | 0.116 | 0.058 | 4F3 |
| 54046 | 45 | Human zinc finger protein ZNF139 | | p0018r11c11 | R0068 E6 | 2.39 | 0.179 | 0.075 | 4E6 |

| Clone | SEQ ID NO. | Genbank | EST | Element (384) | Element (96) | Ratio | Median Signal 1 | Median Signal 2 | 96 Well Location |
|-------|------------|---|-----------------------|---------------|--------------|-------|-----------------|-----------------|------------------|
| 54047 | 46 | Gene for membrane cofactor protein | | p0018r11c16 | R0068 F8 | 3.09 | 0.196 | 0.063 | 4E 7 |
| 54040 | 47 | Colon Kruppel-like factor | | p0018r09c08 | R0068 B4 | 5.44 | 0.377 | 0.069 | 4D12 |
| 54035 | 48 | Human capping protein alpha subunit isoform 1 | | p0018r08c16 | R0067 H8 | 2.17 | 0.157 | 0.072 | 4D7 |
| 54130 | 49 | Ig lambda-chain | | p0021r02c19 | R0078 C10 | 2.41 | 0.076 | 0.032 | 5D6 |
| 54045 | 50 | Protein tyrosine kinase | Placenta/Liver/testis | p0018r10c22 | R0068 D11 | 2.15 | 0.148 | 0.069 | 4E 5 |
| 54050 | 52 | Human microtubule-associated protein 7 | | p0018r11c24 | R0068 F12 | 2.51 | 0.171 | 0.068 | 4E 10 |
| 54051 | 53 | Human retinoblastoma susceptibility protein | | p0018r12c20 | R0068 H10 | 2.02 | 0.172 | 0.085 | 4E 11 |
| 54178 | 54 | Human reticulocalbin | | p0023r06c09 | R0087 C5 | 2.02 | 0.127 | 0.063 | 5H6 |

| Clone | SEQ ID NO. | Genbank | EST | Element (384) | Element (96) | Ratio | Median Signal 1 | Median Signal 2 | 96 Well Location |
|-------|------------|--|-----|---------------|--------------|-------|-----------------|-----------------|------------------|
| 54148 | 55 | Translation initiation factor eIF3 p36 subunit | | p0021r13c01 | R0081 A1 | 2.67 | 0.18 | 0.067 | 5E 12 |
| 54058 | 56 | Human apurinic/apyrimidinic endonuclease | | p0018r13c12 | R0069 B6 | 2.31 | 0.105 | 0.045 | 4F6 |
| 54126 | 58 | Human integral transmembrane protein 1 | | p0021r01c05 | R0078 A3 | 2.31 | 0.117 | 0.051 | 5D2 |
| 54127 | 59 | Human serine kinase | | p0021r01c15 | R0078 A8 | 2.31 | 0.171 | 0.074 | 5D3 |
| 54049 | 60 | Human CGI-44 protein | | p0018r11c18 | R0068 F9 | 2.24 | 0.191 | 0.085 | 4E 9 |
| 54056 | 61 | HADH/NADPH thyroid oxidase p138-tox protein | | p0018r13c05 | R0069 A3 | 2.41 | 0.149 | 0.062 | 4F4 |
| 54064 | 62 | Human peptide transporter (TAP1) protein | | p0018r15c13 | R0069 E7 | 2.96 | 0.104 | 0.035 | 4F12 |

| Clone | SEQ ID NO. | Genbank | EST | Element (384) | Element (96) | Ratio | Median Signal 1 | Median Signal 2 | 96 Well Location |
|-------|------------|--|-----|---------------|--------------|-------|-----------------|-----------------|------------------|
| 54063 | 64 | Transforming growth factor-beta induced gene product | | p0018r15c10 | R0069 F5 | 3.89 | 0.298 | 0.077 | 4F11 |
| 54119 | 66 | Cytokeratin 8 | | p0020r11c02 | R0076 F1 | 5.56 | 0.193 | 0.035 | 5C7 |
| 54111 | 67 | Human coat protein gamma-cop | | p0020r07c24 | R0075 F12 | 2.05 | 0.076 | 0.037 | 5B11 |
| 54121 | 68 | Bumetanide-sensitive Na-K-Cl cotransporter | | p0020r11c20 | R0076 F10 | 3.76 | 0.358 | 0.095 | 5C9 |
| 54125 | 71 | Autoantigen calreticulin | | p0020r16c20 | R0077 H10 | 2.09 | 0.16 | 0.076 | 5D1 |
| 54143 | 72 | Human hepatic squalene synthetase | | p0021r09c21 | R0080 A11 | 2.16 | 0.132 | 0.061 | 5E7 |
| 54139 | 73 | Human RAD21 homolog | | p0021r05c12 | R0079 B6 | 2.26 | 0.06 | 0.026 | 5E3 |
| 54137 | 74 | Human MHC class II HLA-DR-alpha | | p0021r05c08 | R0079 B4 | 2.16 | 0.097 | 0.045 | 5E1 |
| 54044 | 75 | Human Claudin-7 | | p0018r10c12 | R0068 D6 | 5.03 | 0.277 | 0.055 | 4E4 |

| Clone | SEQ ID NO. | Genbank | EST | Element (384) | Element (96) | Ratio | Median Signal 1 | Median Signal 2 | 96 Well Location |
|-------|------------|-------------------------------------|------------------|---------------|--------------|-------|-----------------|-----------------|------------------|
| 54042 | 76 | Ribosome protein S6 kinase 1 | | p0018r09c20 | R0068 B10 | 3.56 | 0.116 | 0.033 | 4E2 |
| 54043 | 77 | CO-029 tumor associated antigen | Colon/Pancreatic | p0018r10c11 | R0068 C6 | 2.65 | 0.18 | 0.068 | 4E3 |
| 54157 | 79 | Human lipocortin II | | p0022r02c18 | R0082 D9 | 3.84 | 0.265 | 0.069 | 5F9 |
| 54116 | 84 | Tumor antigen L6 | | p0020r10c03 | R0076 C2 | 2 | 0.105 | 0.052 | 5C4 |
| 54151 | 85 | UDP-N-acetylglucosamine transporter | | p0021r14c15 | R0081 C8 | 2.35 | 0.093 | 0.04 | 5F3 |
| 54115 | 87 | Cystine/glutamate transporter | | p0020r09c16 | R0076 B8 | 2.05 | 0.033 | 0.016 | 5C3 |
| 54155 | 89 | GAPDH | | p0022r01c04 | R0082 B2 | 4.23 | 0.417 | 0.099 | 5F7 |
| 54169 | 92 | Neutrophil lipocalin | | p0022r15c24 | R0085 F12 | 2.74 | 0.216 | 0.079 | 5G9 |
| 54167 | 93 | Nuclear matrix protein NRP/B | | p0022r13c20 | R0085 B10 | 2.38 | 0.084 | 0.035 | 5G7 |
| 54163 | 97 | Poly A binding protein | | p0022r06c14 | R0083 D7 | 3.28 | 0.262 | 0.08 | 5G3 |

| Clone | SEQ ID NO. | Genbank | EST | Element (384) | Element (96) | Ratio | Median Signal 1 | Median Signal 2 | 96 Well Location |
|-------|------------|---|-----|---------------|--------------|-------|-----------------|-----------------|------------------|
| 54164 | 98 | Ribosome protein L13 | | p0022r08c13 | R0083 G7 | 2.01 | 0.105 | 0.052 | 5G4 |
| 54132 | 99 | Human alpha enolase | | p0021r03c13 | R0078 E7 | 2.96 | 0.292 | 0.099 | 5D8 |
| 54112 | 100 | Human E-1 enzyme | | p0020r08c03 | -R0075 G2 | 2.06 | 0.097 | 0.047 | 5B12 |
| 54165 | 102 | Human ZW10 interactor Zwiint | | p0022r09c22 | R0084 B11 | 2.46 | 0.055 | 0.022 | 5G5 |
| 54158 | 103 | Bumetanide-sensitive Na-K-Cl cotransporter | | p0022r03c20 | R0082 F10 | 2.61 | 0.241 | 0.092 | 5F10 |
| 54108 | 108 | NADH-ubiquinone oxidoreductase NDUF52 subunit | | p0020r06c11 | R0075 C6 | 2.07 | 0.105 | 0.051 | 5B8 |
| 54175 | 109 | Phospholipase A2 | | p0023r04c03 | R0086 G2 | 3.28 | 0.187 | 0.057 | 5H3 |
| 54177 | 111 | Ig heavy chain variable region | | p0023r05c08 | R0087 B4 | 2.31 | 0.117 | 0.051 | 5H5 |
| 54170 | 112 | Protein phosphatase 2C gamma | | p0022r16c04 | R0085 H2 | 2.03 | 0.136 | 0.067 | 5G10 |
| 54176 | 113 | Cyclin protein | | p0023r04c06 | R0086 H3 | 2.12 | 0.165 | 0.078 | 5H4 |

| Clone | SEQ ID NO. | Genbank | EST | Element (384) | Element (96) | Ratio | Median Signal 1 | Median Signal 2 | 96 Well Location |
|-------|------------|------------------------------------|-----|---------------|--------------|-------|-----------------|-----------------|------------------|
| 54180 | 114 | Transgelin 2 (predicted) | | p0023r09c09 | R0088 A5 | 2.21 | 0.166 | 0.075 | 5H8 |
| 53897 | 115 | Human GalNAc-T3 gene | | p0016r01c11 | R0058 A6 | 2.46 | 0.179 | 0.073 | 3A1 |
| 54107 | 118 | Alpha topoisomerase truncated form | | p0020r05c22 | R0075 B11 | 2.64 | 0.108 | 0.041 | 5B7 |
| 53902 | 120 | AD022 protein | | p0016r03c04 | R0058 F2 | 2.3 | 0.123 | 0.053 | 3A6 |
| 54004 | 127 | Cytochrome P450 IIIA4 82% | | p0017r12c21 | R0064 G11 | 2.07 | 0.134 | 0.065 | 4A12 |
| 53913 | 128 | CEA | | p0016r05c23 | R0059 A12 | 5.48 | 0.338 | 0.062 | 3B5 |
| 54134 | 129 | Protein phosphatase (KAP1) | | p0021r04c05 | R0078 G3 | 2.05 | 0.138 | 0.067 | 5D10 |
| 53938 | 131 | Alpha enolase | | p0016r12c15 | R0060 G8 | 3.04 | 0.299 | 0.098 | 3D6 |
| 53939 | 132 | Histone deacetylase HD1 | | p0016r12c23 | R0060 G12 | 2.37 | 0.17 | 0.072 | 3D7 |
| 53914 | 134 | Human squalene epoxidase | | p0016r06c03 | R0059 C2 | 2.12 | 0.07 | 0.033 | 3B6 |

| Clone | SEQ ID NO. | Genbank | EST | Element (384) | Element (96) | Ratio | Median Signal 1 | Median Signal 2 | 96 Well Location |
|-------|------------|--|-----|---------------|--------------|-------|-----------------|-----------------|------------------|
| 53915 | 135 | Human aspartyl-tRNA-synthetase alpha-2 subunit | | p0016r06c09 | R0059 C5 | 2.02 | 0.121 | 0.06 | 3B7 |
| 54101 | 136 | Gamma-actin | | p0020r02c20 | R0074 D10 | 2.91 | 0.21 | 0.072 | 5B1 |
| 53922 | 137 | Human AP-mu chain family member mu1B | | p0016r07c21 | R0059 E11 | 2.07 | 0.161 | 0.078 | 3C2 |
| 54023 | 138 | Human Cctg mRNA for chaperonin | | p0018r02c21 | R0066 C11 | 2.87 | 0.192 | 0.067 | 4C7 |
| 53921 | 140 | Human MEGF7 | | p0016r07c20 | R0059 F10 | 2.5 | 0.109 | 0.044 | 3C1 |
| 54002 | 141 | Connexin 26 | | p0017r12c15 | R0064 G8 | 2.13 | 0.133 | 0.063 | 4A10 |
| 54003 | 142 | Human dipeptidyl peptidase IV | | p0017r12c16 | R0064 H8 | 2 | 0.081 | 0.04 | 4A11 |
| 53925 | 144 | Human 2-oxoglutarate dehydrogenase | | p0016r08c16 | R0059 H8 | 2.7 | 0.167 | 0.062 | 3C5 |

| Clone | SEQ ID NO. | Genbank | EST | Element (384) | Element (96) | Ratio | Median Signal 1 | Median Signal 2 | 96 Well Location |
|-------|------------|---|--------------|---------------|--------------|-------|-----------------|-----------------|------------------|
| 53927 | 145 | Rho guanine nucleotide-exchange factor | | p0016r09c12 | R0060 B6 | 2.13 | 0.194 | 0.091 | 3C7 |
| 53937 | 147 | Human colon mucosa-associated mRNA | Normal colon | p0016r11c23 | R0060 E12 | 2.89 | 0.153 | 0.053 | 3D5 |
| 53919 | 151 | Human embryonic lung protein | | p0016r07c16 | R0059 F8 | 2.19 | 0.153 | 0.07 | 3B11 |
| 53972 | 153 | Human leukocyte surface protein CD31 | | p0017r04c18 | R0062 H9 | 2.08 | 0.052 | 0.025 | 3G4 |
| 54144 | 156 | Poly A binding protein | | p0021r09c24 | R0080 B12 | 2.99 | 0.163 | 0.055 | 5E 8 |
| 53929 | 158 | Cystic fibrosis transmembrane conductance regulator | | p0016r10c02 | R0060 D1 | 4.15 | 0.181 | 0.044 | 3C9 |
| 54099 | 163 | Human set gene | | p0020r02c07 | R0074 C4 | 2.19 | 0.133 | 0.061 | 5A11 |
| 53943 | 164 | Human pleckstrin 2 | | p0016r14c15 | R0061 C8 | 3 | 0.155 | 0.052 | 3D11 |
| 54100 | 166 | Tis11d gene | | p0020r02c09 | R0074 C5 | 2.2 | 0.183 | 0.083 | 5A12 |

| Clone | SEQ ID NO. | Genbank | EST | Element (384) | Element (96) | Ratio | Median Signal 1 | Median Signal 2 | 96 Well Location |
|-------|------------|------------------------------------|-----|---------------|--------------|-------|-----------------|-----------------|------------------|
| 53940 | 167 | Cytokine (GRO-gamma) | | p0016r13c17 | R0061 A9 | 2.37 | 0.183 | 0.077 | 3D8 |
| 53941 | 168 | Human p85Mcm mRNA | | p0016r13c23 | R0061 A12 | 2.25 | 0.09 | 0.04 | 3D9 |
| 54007 | 170 | SOX9 | | p0017r13c19 | R0065 A10 | 2.32 | 0.228 | 0.098 | 4B3 |
| 53950 | 171 | VAV-like protein | | p0016r15c14 | R0061 F7 | 2.41 | 0.064 | 0.026 | 3E6 |
| 53968 | 172 | NF-E2 related factor 3 | | p0017r04c10 | R0062 H5 | 2.19 | 0.1 | 0.046 | 3F12 |
| 54092 | 176 | Human argininosuccinate synthetase | | p0019r15c10 | R0073 F5 | 2.73 | 0.199 | 0.073 | 5A4 |
| 54095 | 177 | Human serine kinase | | p0019r16c14 | R0073 H7 | 2.57 | 0.126 | 0.049 | 5A7 |
| 53967 | 179 | Human phospholipase C beta 4 | | p0017r04c08 | R0062 H4 | 2.87 | 0.182 | 0.063 | 3F11 |
| 54032 | 181 | VAV-3 protein | | p0018r08c01 | R0067 G1 | 2.16 | 0.096 | 0.044 | 4D4 |
| 54135 | 184 | Calcium-binding protein S100P | | p0021r04c13 | R0078 G7 | 5.65 | 0.474 | 0.084 | 5D11 |
| 53969 | 185 | Human leupaxin | | p0017r04c14 | R0062 H7 | 2.12 | 0.042 | 0.02 | 3G1 |

| Clone | SEQ ID NO. | Genbank | EST | Element (384) | Element (96) | Ratio | Median Signal 1 | Median Signal 2 | 96 Well Location |
|-------|------------|---|-----|---------------|--------------|-------|-----------------|-----------------|------------------|
| 53970 | 186 | VAV-3 protein | | p0017r04c15 | R0062 G8 | 2.9 | 0.123 | 0.042 | 3G2 |
| 53995 | 188 | hnRNP type A/B protein | | p0017r11c23 | R0064 E12 | 2.31 | 0.106 | 0.046 | 4A3 |
| 54075 | 189 | Human cell cycle control gene CDC2 | | p0019r04c06 | R0070 H3 | 3.57 | 0.222 | 0.062 | 4G11 |
| 54096 | 195 | Human glutaminyl-tRNA synthetase | | p0019r16c15 | R0073 G8 | 2.17 | 0.206 | 0.095 | 5A8 |
| 54110 | 196 | Human 26S proteasome-associated pad 1 homolog | | p0020r07c22 | R0075 F11 | 2.37 | 0.187 | 0.079 | 5B10 |
| 53920 | 197 | Human squalene epoxidase | | p0016r07c18 | R0059 F9 | 3 | 0.205 | 0.068 | 3B12 |
| 53979 | 198 | Human nuclear chloride ion channel protein | | p0017r05c16 | R0063 B8 | 2.2 | 0.116 | 0.053 | 3G11 |
| 53986 | 201 | Human ephrin | | p0017r08c09 | R0063 G5 | 2.15 | 0.212 | 0.099 | 3H6 |
| 53985 | 205 | CD9 antigen | | p0017r08c06 | R0063 H3 | 3.2 | 0.315 | 0.099 | 3H5 |
| 54012 | 207 | Cyclin B | | p0017r14c19 | R0065 C10 | 2.73 | 0.156 | 0.057 | 4B8 |

| Clone | SEQ ID NO. | Genbank | EST | Element (384) | Element (96) | Ratio | Median Signal 1 | Median Signal 2 | 96 Well Location |
|-------|------------|------------------------------|-----|---------------|--------------|-------|-----------------|-----------------|------------------|
| 53990 | 210 | Colon mucosa-associated mRNA | | p0017r09c22 | R0064 B11 | 2.27 | 0.116 | 0.051 | 3H10 |

EXAMPLE 2

C907P IS OVEREXPRESSED IN COLON TUMORS

Using the C907P cDNA sequence, which was discovered from the subtracted cDNA library and cDNA microarray discussed above, the Genbank database was searched. C907P matches with a known gene named Epiregulin (Genbank accession number D30783). Two gene-specific primers were synthesized, and used for PCR amplification to clone this gene from colon cDNAs. The amplified PCR product was sequenced to confirm its identity. Thus, C907P-Epiregulin is a gene up-regulated in colon cancer. PCR was performed under conditions of denaturing cDNA at 94°C for 1 minute, then 35 cycles of 94°C for 30 seconds, 60°C for 30 seconds, 72°C for 2 minutes. Proof-reading polymerase was used for the amplification. The cDNA templates used for the PCR were synthesized from colon tumor mRNA. The amplified products were cloned into the TA cloning vector and the sequences were determined. The C907P DNA sequence is shown in SEQ ID NO:234, and the amino acid sequence is shown in SEQ ID NO:235.

EXAMPLE 3

FULL LENGTH PCR AMPLIFICATION AND CDNA CLONING OF THE C915P COLON TUMOR ANTIGEN

The C915P cDNA sequence (SEQ ID NO:33; also referred to by clone identifier number 54160), discovered from the subtracted cDNA library and cDNA microarray discussed in Example 1, was used to search the Genbank database. C915P was found to have some degree of similarity to a known gene named superoxidegenerating oxidase Mox1 (Genbank accession number AF127763). Two gene-specific primers were designed according to the sequence deposited in Genbank in order to amplify the full-length cDNA. PCR was performed under conditions of denaturing cDNA at 94°C for 1 minute, then 35 cycles of 94°C for 30 second, 60°C for 30 second, 72°C for 2 minutes. Proofreading polymerase was used for the amplification. The cDNA templates used for the PCR were synthesized from colon tumor mRNA. The amplified products were cloned into the TA cloning vector (Invitrogen, Carlsbad, CA) and random clones sequenced by automatic DNA

sequencing to confirm identity. The full-length cDNA and amino acid sequence of C915P is set forth in SEQ ID NO:244 and 245, respectively.

Expression levels of C915P cDNA were further analyzed by real-time PCR. Using this analysis, C915P was confirmed to be overexpressed in colon tumors as compared to a panel of normal tissues. Moderate levels of expression were observed in normal colon tissues. Real-time PCR (see Gibson et al., *Genome Research* 6:995-1001, 1996; Heid et al., *Genome Research* 6:986-994, 1996) is a technique that evaluates the level of PCR product accumulation during amplification. This technique permits quantitative evaluation of mRNA levels in multiple samples. Briefly, mRNA was extracted from colon tumor and normal tissue and cDNA was prepared using standard techniques. Real-time PCR was performed using a Perkin Elmer/Applied Biosystems (Foster City, CA) 7700 Prism instrument. Matching primers and a fluorescent probe were designed for C915P using the primer express program provided by Perkin Elmer/Applied Biosystems (Foster City, CA). Optimal concentrations of primers and probe were initially determined and control (e.g., β -actin) primers and probe were obtained commercially. To quantitate the amount of specific RNA in a sample, a standard curve was generated using a plasmid containing the C915P cDNA. Standard curves were generated using the Ct values determined in the real-time PCR, which are related to the initial cDNA concentration used in the assay. Standard dilutions ranging from 10^{-10} to 10^{-6} copies of the C915P were generally sufficient. In addition, a standard curve was generated for the control sequence. This permitted standardization of initial RNA content of the tissue samples to the amount of control for comparison purposes.

25

EXAMPLE 4

PRODUCTION OF RA12-C915P-F3 RECOMBINANT PROTEIN IN *E. COLI*

C915P (also referred to as clone identifier 54160, and set forth in SEQ ID NOs:33 and 244 (cDNA), and 245 (amino acid)) has 6 transmembrane domains (TMs) with 3 extracellular loops (ED1, ED2, and ED3). The deletion recombinant protein, Ra12-C915P-f3 (set forth in SEQ ID NOs:236 (cDNA) and 237 (amino acid)),

30

is an N-terminal Ra12 fusion of recombinant, modified C915P in pCRX1 vector (EcoR I, Xho I).

Cloning Strategy for Ra12-C915P-f3:

Three sets of primers were designed that were used sequentially to delete two internal transmembrane domains and amplify a recombined internal region of C915P that was cut with EcoRI and XhoI and ligated in frame with Ra12 in the pCRX1 vector.

PCR#1 used primers AW157 and AW156 (SEQ ID NO:241 and 240, respectively) to amplify the entire construct, deleting TM4 - ID3 - TM5. The PCR product (C915P(minusTM4-ID3-TM5) PCR Blunt II TOPO backbone) was purified from agarose gel, ligated by T4 DNA Ligase and transformed into NovaBlue *E. coli* cells with the following standard protocol: the competent *E. coli* cells were thawed on ice, DNA (or ligation mixture) was added, the reaction mixed and incubated on ice for 5 minutes. The *E. coli* cells were heat-shocked at 42°C for 30 seconds, and left on ice for 2 minutes. Enriched growth media was added to the *E. coli* and they were grown at 37°C for 1 hour. The culture was plated on LB (plus appropriate antibiotics) and grown overnight at 37°C. The next day, several colonies were randomly selected for miniprep (Promega, Madison, WI) and were confirmed by DNA sequencing for correctly deleted region. This step was then repeated on a second region of C915P as described below.

PCR#2 used primers AW155 and AW154 (SEQ ID NOs:239 and 238, respectively) to delete TM2, using a confirmed clone from PCR#1 as template. The PCR product (C915P(minusTM2 / TM4-ID3-TM5) PCR Blunt II TOPO backbone) was purified, ligated and transformed using standard protocols into NovaBlue cells, yielding clones that were confirmed by sequencing for the correct deletion.

PCR#3 used primers AW158 and AW159 (SEQ ID NOs:242 and 243, respectively) to amplify the deleted, recombined three-part fusion protein of C915P, ED1 - ID2-TM3-ED2 - ED3, using the confirmed PCR#2 clone as template. PCR product from PCR#3 was purified and digested using EcoR I and Xho I for ligation into the pCRX1 vector (EcoR I, Xho I). The ligation mixture was transformed into NovaBlue cells by standard protocols, and several clones were selected for miniprep

and sequencing. UI#70526 was confirmed by DNA sequencing to be the correct pCRX1 Ra12-C915P-f3 construct.

Cloning Primers:

5 C915P-AW154 (SEQ ID NO:238): antisense cloning primer to delete TM2, 5'P—Primer Id9682: 5' P- TTTTCTTGTGTAGTAGTATTTGTCG.

C915P-AW155 (SEQ ID NO:239): sense cloning primer to delete TM2, 5'P—Id 9683: 5' P-TGTCGCAATCTGCTGTCCTTCC.

10 C915P-AW156 (SEQ ID NO:240): antisense cloning primer to delete TM4-TM5 region, 5'-P, --Primer Id 9684: 5' P- GCTGGTGAATGTCACATACTCC.

C915P-AW157 (SEQ ID NO:241): sense cloning primer to delete TM4-TM5 region, 5'-P — Id 9685: 5' P- CGGGGTCAAACAGAGGAGAG.

15 Ra12-C915P-F3-AW158 (SEQ ID NO:242): sense cloning primer for the fusion protein with EcoR I site Primer Id 9686: 5' gtcgaattcGATGCCTTCCTGAAATATGAGAAG.

Ra12-C915P-F3-AW159 (SEQ ID NO:243): antisense cloning primer for the fusion protein with stop and Xho I site — Primer Id 9687: 5' cacctcgagttaAGACTCAGGGGGGATGCCCTTC.

Protein Information for Ra12-C915P-f3:

20

Molecular Weight 32429.45 Daltons

297 Amino Acids

28 Strongly Basic(+) Amino Acids (K,R)

27 Strongly Acidic(-) Amino Acids (D,E)

93 Hydrophobic Amino Acids (A,I,L,F,W,V)

25

86 Polar Amino Acids (N,C,Q,S,T,Y)

7.776 Isoelectric Point

3.711 Charge at PH 7.0

Protein Expression:

30 Mini expression screens were performed to determine the optimal induction conditions for Ra12-C915P-f3. The best *E. coli* strain/culture conditions

were screened by transforming the expression construct into different hosts, then varying temperature, culture media and/or IPTG concentration after the inducer IPTG was added to the mid-log phase culture. The recombinant protein expression was then analyzed by SDS-PAGE and/or Western blot. *E. coli* expression hosts BLR (DE3) and
5 HMS (DE3) (Novagen, Madison, WI) were tested in various culture conditions, with little full-length Ra12-C915P-f3 expression detected and Western blots showing some bands at unexpected molecular weights. Tuner (DE3) cells (Novagen, Madison, WI) were then tested with helper plasmids at various IPTG concentrations. Coomassie stained SDS-PAGE showed no induced band but Western blot confirmed a strong
10 Ra12-C915P-f3 signal at 32kD probing with an anti-6xhis tag antibody. The most optimal expression for pCRX1 Ra12-C915P-f3 was found to be in the host strain Tuner (DE3) with a helper plasmid grown in Soy Terrific Broth media at 37°C induced with 1.0 mM IPTG at 37°C for 3hr.

15

EXAMPLE 5

PURIFICATION OF RA12-C915P-F3 RECOMBINANT FUSION PROTEIN FROM *E. COLI*

The clone C915P was found to be over-expressed in a majority of colon cancer tissues. For expression in *E. coli*, the construct Ra12-C915P-f3 (SEQ ID NO:236) was made as described in Example 4. This construct encodes a fusion protein
20 consisting of an N-terminal 6x histidine tag followed by Ra12 and modified C915P (excluding 5 of 6 transmembrane domains) (SEQ ID NO:237). The 32.4kD protein was expressed in multiple large baffled shaker flasks containing 1L of Soy Terrific Broth media. The cultures were spun and cell pellets washed, respun and frozen for purification. After cell lysis, the recombinant protein was found in the insoluble
25 inclusion body fraction. The inclusion body was thoroughly washed with buffered detergents multiple times, then the protein pellet was denatured, reduced and solubilized in buffered 8M Urea and Ra12-C915P-f3 protein was bound to a Ni-NTA affinity chromatography matrix. The matrix was washed to rinse away contaminating
30 *E. coli* proteins and Ra12-C915P-f3 was subsequently eluted using high Imidazole concentration. The fractions containing Ra12-C915P-f3 were pooled and slowly dialyzed to allow for renaturation of the protein. The purified Ra12-C915P-f3 was then

filtered and quantified. SDS-PAGE analysis showed the elution pattern off the nickel column with the major band running at the expected weight of about 32kD. This was further confirmed by western blot using an anti-6x His tag antibody. The western blot also revealed evidence of dimers and tetramers of the recombinant. N-terminal
5 sequencing confirmed purity of about 90%. Purified yield was about 2.5 mg/L induction.

Following is a detailed protocol of the production of purified Ra12-C915P-f3.

For the frozen bacterial cell pellet:

- 10 1. Thaw bacterial cell pellet from 1L induction on ice
2. Add 25ml sonication buffer (20mM Tris, 500mM NaCl) per liter of induction culture
3. Add 1 Complete protease inhibitor tablet and 2mM PMSF (Phenylmethylsulfonyl fluoride) to sonication buffer/pellet mix
- 15 4. Completely resuspend pellet with pipet
5. Add 0.5mg/ml lysozyme (made fresh from lyophilized lysozyme stored at -20°C)
6. Decant into a glass beaker + stir bar, gently stir at 4°C, 30 min
7. French Press 2 x 1100psi, keep on ice
- 20 8. Once lysis solution** has low viscosity, spin at 11000RPM, 30min, 4°C
9. Save supernatant** and pellet

For the pellet from step 9 above:

- 25 1. Wash pellet with 25ml 0.5% CHAPS (3-([3-Cholamidopropyl]dimethylammonio)-1-propanesulfonate) wash (20mM Tris (8.0), 500mM NaCl) ** by sonicating 2x15sec @15Watt
2. Spin at 11000RPM for 25min. Repeat 5x**
- 30 3. Repeat above steps 3 times with 0.5% DOC (Deoxycholic Acid) wash (20mM Tris (8.0), 500mM NaCl)

5

10

15

20

25

4. Resuspend pellet in pellet binding buffer (20mM Tris (8.0), 500mM NaCl, 8M Urea, 20mM Imidazole, 10mM β -Mercaptoethanol) with sonication
5. Equilibrate Ni ++ NTA (Nitrilotriacetic acid) resin (Qiagen, Valencia, CA) with pellet binding buffer, spin down and decant wash (use 4ml resin)
6. Add resin to resuspended pellet, stir at room temperature for 45min
7. Prepare column and buffers, rinse column with pellet binding buffer
8. Pour pellet/Ni resin into column, collect flow through (FT)**
9. Wash column with 30ml pellet binding buffer **
10. Wash column with 30ml pellet binding buffer with 0.5% DOC (Deoxycholic Acid)**
11. Wash column with 30ml pellet binding buffer
12. Elute with 5 x 5ml fractions of pellet binding buffer #1 (binding buffer +300mM Imidazole)**
13. Elute with 2 x 5ml fractions of pellet elution buffer #2 (binding buffer +300mM Imidazole, pH 4.5)**
14. Run SDS-PAGE to screen purification steps (western and coomassie stain)

**Save an aliquot at 4°C for each purification step to check on SDS-PAGE.

EXAMPLE 6

REAL-TIME PCR ANALYSIS OF COLON TUMOR CANDIDATE GENES

The first-strand cDNA to be used in the quantitative real-time PCR was synthesized from 20 μ g of total RNA that had been treated with DNase I (Amplification Grade, Gibco BRL Life Technology, Gaithersburg, MD), using Superscript Reverse Transcriptase (RT) (Gibco BRL Life Technology, Gaithersburg, MD). Real-time PCR

was performed with a GeneAmp™ 5700 sequence detection system (PE Biosystems, Foster City, CA). The 5700 system uses SYBR™ green, a fluorescent dye that only intercalates into double stranded DNA, and a set of gene-specific forward and reverse primers. The increase in fluorescence is monitored during the whole amplification process. The optimal concentration of primers was determined using a checkerboard approach and a pool of cDNAs from breast tumors was used in this process. The PCR reaction was performed in 25µl volumes that include 2.5µl of SYBR green buffer, 2µl of cDNA template and 2.5µl each of the forward and reverse primers for the gene of interest. The cDNAs used for RT reactions were diluted 1:10 for each gene of interest and 1:100 for the β-actin control. In order to quantitate the amount of specific cDNA (and hence initial mRNA) in the sample, a standard curve is generated for each run using the plasmid DNA containing the gene of interest. Standard curves were generated using the Ct values determined in the real-time PCR which were related to the initial cDNA concentration used in the assay. Standard dilution ranging from 20-2x10⁶ copies of the gene of interest was used for this purpose. In addition, a standard curve was generated for β-actin ranging from 200fg-2000fg. This enabled standardization of the initial RNA content of a tissue sample to the amount of β-actin for comparison purposes. The mean copy number for each group of tissues tested was normalized to a constant amount of β-actin, allowing the evaluation of the over-expression levels seen with each of the genes.

Colon tumor candidate genes, C906P (SEQ ID NO:5), C907P (SEQ ID NO:234 (cDNA) and 235 (amino acid)), C911P (SEQ ID NO:21), C915P (SEQ ID NO:244 (cDNA) and 245 (amino acid)), C943P (SEQ ID NO:140), and C961P (SEQ ID NO:200), were analyzed by real-time PCR, as described above, using the short and extended colon panel. These genes were found to have increased mRNA expression in 30-50% of colon tumors. For C906P, slightly elevated expression was also observed in normal trachea, heart, and normal colon. For C907P, elevated expression was also observed in activated PBMC and slightly elevated expression in heart and normal colon. For C911P, slightly elevated expression was observed in pancreas. For C915P, no expression was observed in normal tissues except normal colon. For C943P, no expression was observed in normal tissues except normal colon. For C961P, some

expression was observed in trachea and normal colon. Collectively, the data indicate that these colon tumor candidate genes could be potential targets for immunotherapy and cancer diagnosis.

EXAMPLE 7

PEPTIDE PRIMING OF T-HELPER LINES

Generation of CD4⁺ T helper lines and identification of peptide epitopes derived from tumor-specific antigens that are capable of being recognized by CD4⁺ T cells in the context of HLA class II molecules, is carried out as follows:

Fifteen-mer peptides overlapping by 10 amino acids, derived from a tumor-specific antigen, are generated using standard procedures. Dendritic cells (DC) are derived from PBMC of a normal donor using GM-CSF and IL-4 by standard protocols. CD4⁺ T cells are generated from the same donor as the DC using MACS beads (Miltenyi Biotec, Auburn, CA) and negative selection. DC are pulsed overnight with pools of the 15-mer peptides, with each peptide at a final concentration of 0.25 µg/ml. Pulsed DC are washed and plated at 1×10^4 cells/well of 96-well V-bottom plates and purified CD4⁺ T cells are added at 1×10^5 /well. Cultures are supplemented with 60 ng/ml IL-6 and 10 ng/ml IL-12 and incubated at 37°C. Cultures are restimulated as above on a weekly basis using DC generated and pulsed as above as antigen presenting cells, supplemented with 5 ng/ml IL-7 and 10 U/ml IL-2. Following 4 *in vitro* stimulation cycles, resulting CD4⁺ T cell lines (each line corresponding to one well) are tested for specific proliferation and cytokine production in response to the stimulating pools of peptide with an irrelevant pool of peptides used as a control.

EXAMPLE 8

GENERATION OF TUMOR-SPECIFIC CTL LINES USING IN VITRO WHOLE-GENE PRIMING

Using *in vitro* whole-gene priming with tumor antigen-vaccinia infected DC (see, for example, Yee et al, *The Journal of Immunology*, 157(9):4079-86, 1996), human CTL lines are derived that specifically recognize autologous fibroblasts transduced with a specific tumor antigen, as determined by interferon-γ ELISPOT

analysis. Specifically, dendritic cells (DC) are differentiated from monocyte cultures derived from PBMC of normal human donors by growing for five days in RPMI medium containing 10% human serum, 50 ng/ml human GM-CSF and 30 ng/ml human IL-4. Following culture, DC are infected overnight with tumor antigen-recombinant vaccinia virus at a multiplicity of infection (M.O.I) of five, and matured overnight by the addition of 3 μ g/ml CD40 ligand. Virus is then inactivated by UV irradiation. CD8+ T cells are isolated using a magnetic bead system, and priming cultures are initiated using standard culture techniques. Cultures are restimulated every 7-10 days using autologous primary fibroblasts retrovirally transduced with previously identified tumor antigens. Following four stimulation cycles, CD8+ T cell lines are identified that specifically produce interferon- γ when stimulated with tumor antigen-transduced autologous fibroblasts. Using a panel of HLA-mismatched B-LCL lines transduced with a vector expressing a tumor antigen, and measuring interferon- γ production by the CTL lines in an ELISPOT assay, the HLA restriction of the CTL lines is determined.

15

EXAMPLE 9

GENERATION AND CHARACTERIZATION OF ANTI-TUMOR ANTIGEN MONOCLONAL ANTIBODIES

20 Mouse monoclonal antibodies are raised against *E. coli* derived tumor antigen proteins as follows: Mice are immunized with Complete Freund's Adjuvant (CFA) containing 50 μ g recombinant tumor protein, followed by a subsequent intraperitoneal boost with Incomplete Freund's Adjuvant (IFA) containing 10 μ g recombinant protein. Three days prior to removal of the spleens, the mice are 25 immunized intravenously with approximately 50 μ g of soluble recombinant protein. The spleen of a mouse with a positive titer to the tumor antigen is removed, and a single-cell suspension made and used for fusion to SP2/O myeloma cells to generate B cell hybridomas. The supernatants from the hybrid clones are tested by ELISA for specificity to recombinant tumor protein, and epitope mapped using peptides that 30 spanned the entire tumor protein sequence. The mAbs are also tested by flow

cytometry for their ability to detect tumor protein on the surface of cells stably transfected with the cDNA encoding the tumor protein.

From the foregoing it will be appreciated that, although specific embodiments of the invention have been described herein for purposes of illustration, 5 various modifications may be made without deviating from the spirit and scope of the invention. Accordingly, the invention is not limited except as by the appended claims.

CLAIMS

What is Claimed:

1. An isolated polynucleotide comprising a sequence selected from the group consisting of:

- (a) sequences provided in SEQ ID NO:1-234, 236, and 244;
- (b) complements of the sequences provided in SEQ ID NO:1-234, 236, and 244;
- (c) sequences consisting of at least 20 contiguous residues of a sequence provided in SEQ ID NO:1-234, 236, and 244;
- (d) sequences that hybridize to a sequence provided in SEQ ID NO:1-234, 236, and 244, under moderately stringent conditions;
- (e) sequences having at least 75% identity to a sequence of SEQ ID NO:1-234, 236, and 244;
- (f) sequences having at least 90% identity to a sequence of SEQ ID NO:1-234, 236, and 244; and
- (g) degenerate variants of a sequence provided in SEQ ID NO:1-234, 236, and 244.

2. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:

- (a) sequences encoded by a polynucleotide of claim 1;
- (b) amino acid sequences set forth in SEQ ID NO:235, 237, and 245;
- (c) sequences having at least 70% identity to a sequence encoded by a polynucleotide of claim 1; and
- (d) sequences having at least 90% identity to a sequence encoded by a polynucleotide of claim 1.

3. An expression vector comprising a polynucleotide of claim 1 operably linked to an expression control sequence.

4. A host cell transformed or transfected with an expression vector according to claim 3.

5. An isolated antibody, or antigen-binding fragment thereof, that specifically binds to a polypeptide of claim 2.

6. A method for detecting the presence of a cancer in a patient, comprising the steps of:

- (a) obtaining a biological sample from the patient;
- (b) contacting the biological sample with a binding agent that binds to a polypeptide of claim 2;
- (c) detecting in the sample an amount of polypeptide that binds to the binding agent; and
- (d) comparing the amount of polypeptide to a predetermined cut-off value and therefrom determining the presence of a cancer in the patient.

7. A fusion protein comprising at least one polypeptide according to claim 2.

8. An oligonucleotide that hybridizes to a sequence recited in SEQ ID NO:1-234, 236, and 244 under moderately stringent conditions.

9. A method for stimulating and/or expanding T cells specific for a tumor protein, comprising contacting T cells with at least one component selected from the group consisting of:

- (a) polypeptides according to claim 2;
- (b) polynucleotides according to claim 1; and
- (c) antigen-presenting cells that express a polypeptide according to claim 1,

under conditions and for a time sufficient to permit the stimulation and/or expansion of T cells.

10. An isolated T cell population, comprising T cells prepared according to the method of claim 9.

11. A composition comprising a first component selected from the group consisting of physiologically acceptable carriers and immunostimulants, and a second component selected from the group consisting of:

- (a) polypeptides according to claim 2;
- (b) polynucleotides according to claim 1;
- (c) antibodies according to claim 5;
- (d) fusion proteins according to claim 7;
- (e) T cell populations according to claim 10; and
- (f) antigen presenting cells that express a polypeptide according to claim 2.

12. A method for stimulating an immune response in a patient, comprising administering to the patient a composition of claim 11.

13. A method for the treatment of a cancer in a patient, comprising administering to the patient a composition of claim 11.

14. A method for determining the presence of a cancer in a patient, comprising the steps of:

- (a) obtaining a biological sample from the patient;
- (b) contacting the biological sample with an oligonucleotide according to claim 8;
- (c) detecting in the sample an amount of a polynucleotide that hybridizes to the oligonucleotide; and
- (d) compare the amount of polynucleotide that hybridizes to the oligonucleotide to a predetermined cut-off value, and therefrom determining the presence of the cancer in the patient.

15. A diagnostic kit comprising at least one oligonucleotide according to claim 8.

16. A diagnostic kit comprising at least one antibody according to claim 5 and a detection reagent, wherein the detection reagent comprises a reporter group.

17. A method for inhibiting the development of a cancer in a patient, comprising the steps of:

(a) incubating CD4+ and/or CD8+ T cells isolated from a patient with at least one component selected from the group consisting of: (i) polypeptides according to claim 2; (ii) polynucleotides according to claim 1; and (iii) antigen presenting cells that express a polypeptide of claim 2, such that T cell proliferate;

(b) administering to the patient an effective amount of the proliferated T cells, and thereby inhibiting the development of a cancer in the patient.

SEQUENCE LISTING

<110> Corixa Corporation
 Jiang, Yuqiu
 Hepler, William T.
 Clapper, Jonathan
 Wang, Aijun
 Secrist, Heather

<120> COMPOSITIONS AND METHODS FOR THE THERAPY
 AND DIAGNOSIS OF COLON CANCER

<130> 210121.524PC

<140> PCT

<141> 2001-06-08

<160> 245

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 334

<212> DNA

<213> Homo sapien

<400> 1

| | | | | | | |
|------------|-------------|------------|------------|------------|-------------|-----|
| actcaatatt | ataaaaaacct | caaataattg | acttgatttt | acacaacatc | cttccctttt | 60 |
| ctacaagtta | atTTTTTTac | aaatcatttg | ggttatctcc | taaatagggt | atattttatt | 120 |
| gcttctagaa | acaatgtttc | aaaatatatg | tgcatatatc | gtaataattt | gtataaatat | 180 |
| ttcccacaac | aattttcata | atTTTcaaag | actaatttct | tgactgaaga | tatttttgcta | 240 |
| gggaagtga | actttaaaat | tttgtagatt | ttaaaaaata | ttgttgaatg | gtgtcatgca | 300 |
| aaggatttat | atagtgtgct | cccactaact | gtgt | | | 334 |

<210> 2

<211> 650

<212> DNA

<213> Homo sapien

<220>

<221> misc_feature

<222> (1)...(650)

<223> n = A,T,C or G

<400> 2

| | | | | | | |
|------------|------------|------------|------------|-------------|------------|-----|
| actttgttat | ttttccatca | ctaaaggcca | atcagaattt | ggaaccatgc | tgctacccaa | 60 |
| gaaatcta | ggaatgaatt | agttctgtag | atgacaattt | cttcacccat | ttatgagacc | 120 |
| taaatctttt | ccataacact | catgtattca | gtataacaac | atactaactg | aaagagggac | 180 |
| ctgattgttt | aaagtttgat | tgacagacgt | gtagaacata | actcattatg | tttcagataa | 240 |
| ggtaactcct | agatatcaaa | ctaatttggt | ggggtagaga | ttttacaagt | catgccatta | 300 |
| gaagattttc | tctgatatta | tatgtgcagt | tcagttacaa | gatgaaatca | tgTTTTTTta | 360 |
| acaaaagaga | taaaatacaa | ttgaagcaaa | aaataacagc | tagtatataa | tatatacagt | 420 |
| ctgtattttg | ttttcacagt | aggcctgatg | actaaaagat | atgctttatt | acacgctatt | 480 |
| ttcacctcct | gaaagtcaaa | ggtgatgatt | aatttcattt | agcaggggaag | tggaataata | 540 |

tcttttgaaa taactaagtc cactaaatta tcagtatgct attctggggt ctaagtacct 600
gnccggcggn cgctcaaang gcgaattctg cagatatnca tcaccttggc 650

<210> 3
<211> 444
<212> DNA
<213> Homo sapien

<400> 3
acacatccca tcttcaaatt taaaatcata ttgtcagttg tccaaagcag cttgaattta 60
aagtttgtgc tataaaattg tgcaaataatg ttaaggattg agaccaccca atgcactact 120
gtaatatctt gcttcctaaa tttcttccac ctacagataa tagacaacaa gtctgagaaa 180
ctaaggctaa ccaaacttag atataaatcc taccaataaa atttttcagt ttttaagtttt 240
acagtttgat ttaaaaacaa aacagaaaca aatttcaaaa taaatcacat cttctcttaa 300
aacttggcaa acccttcctt aactgtccaa gtatgagcat acactgccac tggctttaga 360
tactccaatt aaatgcacta ctctttcact ggtctgaatg aagtatgggtg aaacaagtac 420
ctgcccgggc gggcaagggc gaat 444

<210> 4
<211> 509
<212> DNA
<213> Homo sapien

<400> 4
aaaaacaaaa ttaaattttc atttcaatta agaccctttt tggcattttg cttatttatt 60
ctgccctttg gttaacagca tcagcatcac attactattt tatattgcat atatgtagca 120
tttgcttctt taagttttca acatatcatt tatattttaa ggcagacact gagtcagtat 180
taatagatta actaaactgc actgtaattt agataaaatt actgtgtctc actgtgtatt 240
acatgcaaaa tccacataaa ttgtcattta accaacagta ctgcacgagc gaacatctcg 300
atatatgaaa actgcatcat caattcaacg ttttggtact tgaaactgca tcataaatgc 360
aacattgtca tatgtgaaaa cgacacctta agtccttctt tttaaaaatg acattgcgtt 420
tagcttattg taagaggttg aactttttgta ttttgtaact atctttaagc tcttcagttt 480
ataattcata taaaatgcct tttgtattt 509

<210> 5
<211> 478
<212> DNA
<213> Homo sapien

<400> 5
acattgagta gagcatcaag agcaataaaa aagacttcaa aaaagggttac aagagcattc 60
tcttttctcca aaactccaaa aagagctctt cgaagggctc ttatgacatc ccacggctca 120
gtggagggaa gaagtccttc cagcaatgat aagcatgtaa tgagtcgtct ttctagcaca 180
tcatcattag caggtatccc ttctccctcc cttgtcagcc ttcttctctt ctttgaaagg 240
agaagtcata cgtttaagtag atctacaact catttgatat gaagcggttac caaatctta 300
aattatagaa atgtatagac acctcactact caaataagaa actgacttaa atggtacctg 360
cccgggcggc caagggcgaa ttctgcagat atccatcaca ctggcggcgc ctcgagcatg 420
catctagagg gcccaattcg ccctatagtg agtcgtatta caattcactg gccgtcgt 478

<210> 6
<211> 485
<212> DNA
<213> Homo sapien

<400> 6
aaatgtccaa ggtggcccca agggaggact tctgcagcac agctcccttc ccaggacgtg 60
aaaatctgcc ttctcaccat gaggttcta gtcctttcca gctgctctg tatectgctt 120
ctctgcttct ccatcttctc cacagaaggg aagaggcgtc ctgccaaggc ctggtcaggc 180

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| aggagaacca | ggctctgctg | ccaccgagtc | cctagcccca | actcaacaaa | cctgaaagga | 240 |
| catcatgtga | ggctctgtaa | accatgcaag | cttgagccag | agccccgcct | ttgggtggtg | 300 |
| cctggggcac | tcccacaggt | gtagcactcc | caaagcaaga | ctccagacag | cggagaacct | 360 |
| catgcctggc | acctgaggtg | cctgcccggg | cggccaaggg | cgaattctgc | agatatccat | 420 |
| cacactggcg | ggccgctcga | gcatgcatct | agagggccca | attcgcccta | tagtgagtcg | 480 |
| tatta | | | | | | 485 |

<210> 7

<211> 483

<212> DNA

<213> Homo sapien

<220>

<221> misc_feature

<222> (1) ... (483)

<223> n = A, T, C or G

<400> 7

| | | | | | | |
|------------|-------------|------------|------------|-------------|------------|-----|
| actgctggct | gccccggctg | gtcagtgggg | caaagccggg | catgaagaag | tgcagccggg | 60 |
| gaaacgggac | catgttcaca | gccagcttcc | gcaggtcagc | attgagctgg | cctgggaagc | 120 |
| gcaggcaggt | ggtgacccca | ctcatggtag | cagacaccag | gtggttcagg | tcaccatagg | 180 |
| tgggcgtggt | cagcttttagg | gttctgaagc | aaatgtcgta | gagagcttcg | ttatcaatgc | 240 |
| agtaggtctc | gtctgtgttt | tctacgagct | ggtggactga | gaggggtggcg | ttgtagggct | 300 |
| ccaccactgt | gtctgacact | ttgggccaag | gcaccacact | aaacgtgttc | atgatcctgt | 360 |
| ctgggtacct | gccccgggcg | tcgaaagggc | gaattctgca | gataccatc | acactggcng | 420 |
| gccgctcgag | catgcatcta | gagggcccaa | ttcgccctat | agtgagtcgt | attacaattc | 480 |
| act | | | | | | 483 |

<210> 8

<211> 398

<212> DNA

<213> Homo sapien

<400> 8

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| acaaggcaga | tggagcattg | acgttttcaa | aaccattatt | cctgtgactg | gagaggcatc | 60 |
| aggagagggc | tcgttcgtct | ccagctcata | aaatgtagca | gcatcatcct | tgacagtgat | 120 |
| gtttttcagg | ccctccattg | agaacctgag | gaaatctgta | aagataagtg | gtgatgttgt | 180 |
| ttcaaacgtt | cagaacagat | accatcatcc | tgcctttgtt | agctgctgta | gggaaagtgc | 240 |
| gttacagatg | tctgctgacc | tcacaagagt | gaaaagataa | actgtgcatg | tgtttccact | 300 |
| tccgtttcta | gtacctgccc | gggcggcaag | ggcgaattct | gcagatatcc | atcacactgg | 360 |
| gcgccgctcg | agcatgcata | tagagggcc | aattcgcc | | | 398 |

<210> 9

<211> 493

<212> DNA

<213> Homo sapien

<400> 9

| | | | | | | |
|------------|------------|------------|------------|-------------|------------|-----|
| acagctttta | tatctggagt | agctatttag | tgtccttct | ctacctaagc | aaggtttgac | 60 |
| tgatagtcac | tggagttttc | ctgcagaact | tggtcatatc | cactcatact | gctctgacca | 120 |
| ccataaccac | ctccataacc | accactcagc | tgttggttag | caggacctcc | ataactagac | 180 |
| tggttggata | agcccatccc | tcccatcatt | tggctaccat | aagcgccacc | acttgcccct | 240 |
| gctgtagaat | tcaaaaaaag | ttctacatag | ctgtgatcgt | aagcaccccc | acttgttcct | 300 |
| gcagtagaat | ttaagaagag | ctccacatat | ctgtgttgca | tattagcttt | gtcttttgcc | 360 |
| atagctgcca | cagcatcttc | atgagtagca | aattcaacat | ctgcctcacc | ggtaactctg | 420 |
| ccatcgggtc | caatttcaat | gtgtacctgc | cggggcggca | agggcggaatt | ctgcagatat | 480 |
| ccattacact | ggc | | | | | 493 |

<210> 10
<211> 392
<212> DNA
<213> Homo sapien

<400> 10
acaaaacaca accgaggagc gtatacagtt gaaaacattt ttgttttgat tggaaggcag 60
attattttat attagtatta aaaatcaaac cctatgtttc tttcagatga atcttccaaa 120
gtggattata ttaagcaggt attagattta ggaaaacott tccatttctt aaagtattat 180
caagtgtcaa gatcagcaag tgtccttaag tcaaacaggt tttttttgtt gttgtttttg 240
ctttgtttcc ttttttagaa agttctagaa aataggaaaa cgaaaaattt cattgagatg 300
agtagtgcac ttaattattt tttaaaaaac tttttaagta cgctgtgaag gcatcaacat 360
ttctggcaat ttctacagaa acaagttgaa gt 392

<210> 11
<211> 525
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(525)
<223> n = A,T,C or G

<400> 11
accacaacac caggcctcag tgaggcaten accaccttct acagcagccc cagatcacca 60
accacaacac tctcacctgc cagtatgaca agcctaggcg tcggtgaaga atccaccacc 120
tcccgtagcc aaccagggtc tactcactca acagtgtnac ctgncagcac caccacgcca 180
ggcctcanng aggaatctac caccgnctac agcangcctg agtgagaaat ntaccacttt 240
ncacagtagc cccagatcac cagccacaac actctcacct gccancacga caagctcagg 300
cgtnagtga gaatccacca cctcccacag ncgaccaggc tcaacgcaca caacagcatt 360
ccctgacagn accaccacnc cnggcctcan tnggcattct acaacttccc acagcaannc 420
cangctnaac ggatacaaca ctgttacctg ccaggaccac cacctcagge cccagtcagg 480
aatcaacaac ttcccacagc agnccagggt caactgacac agcac 525

<210> 12
<211> 498
<212> DNA
<213> Homo sapien

<400> 12
accacagcct tatcctttgg tcaagaatct acaaccttcc acagcagccc aggtccact 60
cacacaacac tcttccctga cagcaccaca agctcaggca tcggtgaagc atctacacgc 120
gtccacagca gcaactggctc accacgcaca acaactgtcc ctgccagctc cacaagccct 180
ggacttcagg gagaatctac caccttccag acccaccag cctcaactca cagcagcct 240
tcacctccta gcaccgcaac agcccctgtt gaagaatcta caacctacca ccgcagccca 300
ggctcgactc caacaacaca ctcccctgcc agctccacaa cttcggggcca cagtgaagaa 360
tcaacaatat tccacagcag cccagatgca agtggaacaa caccctcatc tgcccactcc 420
acaacctcag gtcgtggaga atctacaacc tcacgcatca gtccaggctc aactgaaata 480
acaacgttac ctggcagt 498

<210> 13
<211> 523
<212> DNA
<213> Homo sapien

<400> 13
accacagcat catcccttgg tccagaatat actaccttcc acagccgccc aggtccact 60


```

gaaacaacac tcttacctga caacaccaca gcctcaggac tccttgaagc atctacgccc 120
gtccacagca gcaccagatc gccacacaca aactgtccc ctgccggctc tacaacccgt 180
cagggagaat ctaccacatt ccatagctgg ccaagctcaa aggacactag gcccgcacct 240
cctactacca catcagcctt tgttaaacta tctacaactt atcacagcag cccgagctca 300
actccaacaa cccacttttc tgccagctcc acaaccttgg gccatagtga ggaatcgaca 360
ccagtccaca gcagcccagt tgcaactgca acaacacccc cacctgcccg ctccgcgacc 420
tcaggccatg ttgaagaatc tacagcctac cacaggagcc cgggctcaac tcaaacaatg 480
cacttcctg aaagctccac aacttcaggc catagtgaag aat 523

```

<210> 14
 <211> 461
 <212> DNA
 <213> Homo sapien

<220>
 <221> misc_feature
 <222> (1) ... (461)
 <223> n = A,T,C or G

```

<400> 14
caggtacaag tcattactcc cccttctccc atatgaacaa gaatttttta acggtcagaa 60
tatattgggc atcaaattaa aaactttttt ttcaaaagtc tacagaatgg atattggagc 120
aaaaattaca aagtgggtca gatacagggt tttaaaaact gcattactga atttaacaaa 180
agtcagacac tagaatcata tatttgctgc ataaaagttg atttgatacc tgggtggtgat 240
tgaatttagt ctcaaagact cataaataaa aatctgactt aagacgtagt cataccagta 300
taccaattct cccatcactt tgactttcgg cagagagatt agagcaaaaa atattcagga 360
gaacagtgga gttacattgn attatgtatg tttaataata tatcaatttt aagggttaagg 420
ttaaggaaat cttaatttta aggntaaacc ttgagtacct c 461

```

<210> 15
 <211> 508
 <212> DNA
 <213> Homo sapien

<220>
 <221> misc_feature
 <222> (1) ... (508)
 <223> n = A,T,C or G

```

<400> 15
cgcgggcgagg taccagtgtg tgttcgtatt tgggcacagg ctttnggggg ccactgcgtt 60
gcagntgaca tgtgcccagg ttacagttca tttgcgactt cgttcctttg gtgcacttgt 120
tcacacaggc cagcttcccg tccaagacat ccacatagta gaactgggta tacccttcgg 180
cagccttctg ggtgcattgc tcctggaagt caaagcccgg agtcaccgat gaatccacga 240
aagtgtcctc ttactatag cacagtatgg ctttctgca ggaatcagga tcaagaagag 300
ttgttctagt ttcatcata atcttggcct ttacaatctc tgccagggtt tcaaacagtt 360
cctcactatc taaagtgtag tctgcctcca ggatgacatc gttcttgacc acgatgctac 420
cgttgagcaa tctccgaatg ttcacccctc tatactgagg aagattgtcg cccttcaaaa 480
cgacatccat ccgattcttg aagagggt 508

```

<210> 16
 <211> 578
 <212> DNA
 <213> Homo sapien

```

<400> 16
acatataaat gaatctggtg ttggggaaac cttcatctga aaccacaga tgtctctggg 60
gcagatcccc actgtcctac cagttgccct agcccagact ctgagctgct caccggagtc 120

```

| | | | | | | |
|------------|------------|------------|------------|-------------|-------------|-----|
| attggaagg | aaaagtggag | aatggcaag | tctagagtct | cagaaactcc | cctggggggtt | 180 |
| tcacctgggc | cctggaggaa | ttcagctcag | cttcttccta | ggtccaagcc | ccccacacct | 240 |
| tttccccaac | cacagagaac | aagagtttgt | tctgttctgg | gggacagaga | aggcgcttcc | 300 |
| caacttcata | ctggcaggag | ggtgaggagg | ttcactgagc | tccccagatc | tcccactgcg | 360 |
| gggagacaga | agcctggact | ctgccccacg | ctgtggccct | ggaggggtccc | ggtttgtcag | 420 |
| ttcttggtgc | tctgtgttcc | cagaggcagg | cggagggtga | agaaaggaac | ctgggatgag | 480 |
| gggtgctggg | tataagcaga | gagggatggg | ttcctgctcc | aagggaccct | ttgcctttct | 540 |
| tctgcccttt | cctaggccca | ggcctggggt | tgtacctt | | | 578 |

<210> 17
<211> 623
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(623)
<223> n = A,T,C or G

| | |
|------------|------------|
| <400> 17 | |
| acacagaagt | ttgaatcaca |
| aaacataatt | accacaataa |
| aacacagtgt | tcaagtatct |
| 60 | |
| tggcagagca | atctgccgca |
| caaactgcaa | attaaattaa |
| ctacacagac | taaaaactat |
| 120 | |
| acagcctacc | atcaacagtt |
| gtgcattata | aaaaggtagt |
| ttctttcctt | ttgttttaag |
| 180 | |
| tcaggaacag | gtagattttt |
| aaaaatatat | atacaagcta |
| acacacacag | ctatcagcac |
| 240 | |
| taatgcccc | ccctcaactt |
| ttcctttttc | ttatagaaaa |
| tggaaagett | acaatacctc |
| 300 | |
| ctccatcaaa | gcggcaggcc |
| tacgagccag | cctgaacagg |
| gtttgccttg | gaaaagatgt |
| 360 | |
| ggcctgaggt | ttagagccgc |
| tttgtgcggg | gatggtggag |
| gctaggggtg | gggtgagaaa |
| 420 | |
| agggagaagg | cggaaggggg |
| acggacagtt | ctttcttttt |
| ctctctagct | tacccttttt |
| 480 | |
| tctaaataag | cccaaattgc |
| atcactcgtc | ttttgctcgg |
| tctttgttga | ttttcttcat |
| 540 | |
| tttcatcctg | cggttctgga |
| accagatctt | gacctgctct |
| cggtgagggt | gagcagtcga |
| 600 | |
| gcccctcgta | cctgccggcg |
| gnc | |
| 623 | |

<210> 18
<211> 477
<212> DNA
<213> Homo sapien

| | |
|-------------|------------|
| <400> 18 | |
| acacaaaagg | gcatagtcct |
| acaaagtgtg | ttatataatt |
| gttttatgtg | tgcaaattga |
| 60 | |
| aatatataag | atggatcagg |
| gatctcagtt | taaggaatcc |
| tgcttctgt | atgatgatgt |
| 120 | |
| cttaattttt | gagattttca |
| tatatgggt | tatagctata |
| tatcaggaca | ggtaaataca |
| 180 | |
| ttataaaatt | ataaccttta |
| taataatttt | tagtataatc |
| acttgtgtga | ctataataaa |
| 240 | |
| ttggcttttag | ttttctttac |
| tcttcacagt | tttaaatagg |
| aactatttta | caagaataac |
| 300 | |
| attgctaggt | agaaaaattt |
| ctgttcagtt | aggagttcct |
| atgttgctgc | tgaaatgagt |
| 360 | |
| catgcacaat | tttaaatctc |
| tgtagtttct | tcataagcta |
| ttttactatc | ttactatttt |
| 420 | |
| ataagccttg | tggtgcagtc |
| aagttttttac | cacattctat |
| agaccttgct | gtacctg |
| 477 | |

<210> 19
<211> 374
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(374)
<223> n = A,T,C or G

<400> 19

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| agaaacttta | gcattggccc | agtagtggt | tctagctcta | aatgtttgcc | ccgccatccc | 60 |
| tttccacagt | atccttcttc | cctcctcccc | tgtctctggc | tgtctcgagc | agtctagaag | 120 |
| agtgcattct | cagcctatga | aacagctggg | tctttggcca | taagaagtaa | agatttgaag | 180 |
| acagaaggaa | gaaactcagg | agtaagcttc | tagaccctt | cagcttctac | acccttctgc | 240 |
| cctctctcca | ttgcctgcac | cccacccag | ccactcaact | cctgcttggt | tttcctttgg | 300 |
| ccataggaag | gtttaccagt | agaatccttg | ctaggttgat | gtgggccata | cattccttta | 360 |
| ataaaccatt | gngt | | | | | 374 |

<210> 20

<211> 207

<212> DNA

<213> Homo sapien

<400> 20

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| acaagtgtgg | cctcatcaag | ccctgccag | ccaactactt | tgcgtttaaa | atctgcagtg | 60 |
| gggccgcaa | cgctgtgggc | cctactatgt | gctttgaaga | ccgcatgac | atgagtcctg | 120 |
| tgaaaaacaa | tgtgggcaga | ggcctaaaca | tcgccctggt | gaatggaacc | acgggagctg | 180 |
| tgctgggaca | gaaggcattt | gacatgt | | | | 207 |

<210> 21

<211> 557

<212> DNA

<213> Homo sapien

<400> 21

| | | | | | | |
|------------|------------|-------------|-------------|------------|------------|-----|
| acaaagaatc | cctagacgcc | atactgagtt | ttaaagttcct | taattcctaa | tttaaggctt | 60 |
| ctagtgaagc | ctcctcacag | taggcttcac | taggccaca | gtgccctag | acctctgaca | 120 |
| atcccaccct | agacagactt | tattgcaaaa | tgcgcctgaa | gaggcagatg | attcccaaga | 180 |
| gaactcacca | aatcaagaca | aatgtcctag | atctctagt | tggtagaact | atgcacctaa | 240 |
| acattgctgc | aaaatgaaca | cactttttaga | caccctgca | gatattctaa | taagtggaga | 300 |
| agactatttt | ttcaacaaac | attttctctt | tcaccctaac | tcctaaacag | cttactgggg | 360 |
| cttctgcaag | acagaaagat | cataattcag | aaggtaacca | tcgttataga | cataaagttt | 420 |
| ctggtcaaaa | gggttatagt | taatgtctctg | cactttttcc | tgcatcttat | gcattacaat | 480 |
| gtctagtttg | ccctctttcc | ctgtgtttgt | gtcataatag | taaaaaatct | cttctgttct | 540 |
| ggggtcatag | cacctcg | | | | | 557 |

<210> 22

<211> 541

<212> DNA

<213> Homo sapien

<400> 22

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| acctaggtgc | tagtctcccc | actaactgag | ggaaaaaggt | tcccaggtgg | ggtcctctgc | 60 |
| ccactttgcc | accacattca | cattccaaat | gggataatgc | ctgaggggcc | aagagtggtc | 120 |
| aggctgccct | ggggtgaatg | tcaccctgat | gaggcccatc | agctcttgcc | cactcagtga | 180 |
| ggccagactt | gtgctctaat | ccactctcct | gtgggtccct | ggcctgtatg | gcttatactg | 240 |
| gggagctggg | cctctgggct | gtccaaaccc | aagggtcaca | ctttgctttt | cctttgttgt | 300 |
| ccccattttc | cctccttgct | ctaagacaaa | acttttccca | gagaagaact | ctttgttgct | 360 |
| cccgtcagc | tgtaattctg | ccttttctac | cttcattcca | tccttcctct | gccagataa | 420 |
| agtccagcag | aaattcctcc | tttctacctc | tctgggactc | tgagacagga | aatcttcaag | 480 |
| gaggagtttt | tccctcccca | ctattcttat | tctcaacccc | cagaggaacc | aaggctgctg | 540 |
| t | | | | | | 541 |

<210> 23

<211> 486

<212> DNA

<213> Homo sapien

<400> 23
acaaaattgt tggaaatttag ctaatagaaa aacatagtaa atatttacia aaacgttgat 60
aacattactc aagtcacaca catataacia tgtagacagg tcttaaciaa gtttaciaat 120
tgaaattatg gagatttccc aaaatgaatc taatagctca ttgctgagca tggttatcaa 180
tataacattt aagatcttgg atcaaagtgt gtccccgagt cttctacaat ccagtcctct 240
tagaaattgg tttctctctt tgggagattc agactcagag gcagccagag gggacaggtc 300
aagagctgaa ataatcacat aactactcta attttcttca ttctattgac tgtgtcaagt 360
tatagacaca gccaaagtgt ttttcttcgg cctctgatga tttgagaaga tgagaacat 420
gagcaatttc tcattgctta aagaaaaact tggcacataa gaggctgagt gtagtagagt 480
atctgt 486

<210> 24
<211> 450
<212> DNA
<213> Homo sapien

<400> 24
actgatacat gctataacag agatgaactt cgaaaacatg ctaagtgaia gaagccaaat 60
ccaaaaacia taaaaacaca tattgtatcc tcaccctttt cgcatttttag tgagcaatca 120
ttgcatatga atgtttatgg gaaaaatcaa tgtgtgctaa atcattgtat tccagtaa at 180
agattggact taaaacttga tacagaagtt gcaaataagt gggattgagt ttgattatta 240
tatagaaaat aattacatga ttcattttaag aataataata tccaccattt attgagcact 300
tactatgagc ctgtgtgcca aacatttcat gcatttctca ttttaattctc acaataatcc 360
tgtgaggtag aagctattag gttgaatcat atgaacttgc caatatatga taatttctaa 420
gagttgggaa tttttgagga tgtgaatggt 450

<210> 25
<211> 638
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(638)
<223> n = A,T,C or G

<400> 25
gcaggtaaac gtagcgcttc cccgacgtct tgtggatgat gttcttgncc taatagtagc 60
gtaagccccg gtcagcttc tcgtagttca tcttgggctt atttttcctc tttccccacc 120
ggcgggccac ctcatcgagg tcggcgagct taaactccca tccgtctcca gtccagctga 180
tgaatgactg gcaggatttg tctgatagca gctccaggag aaactgccac agctgaatag 240
gtccacttcc tgtgaagccg gccagcacag ctgcagggtat aactggtttg ccttgctcca 300
ccgggtcact cctctcttgg atgtaatcct tgaaagacat ggttggctta ttgaggcaga 360
gagactggct gcagtcactc tcgaagctct cgaaggaagg aaccggtgc acatccagca 420
aggacgactg gctgttccag gactggagga gggagtctga gctctcgagg ctgtccgcac 480
cgttctcagg ggagtcgtgg tctttggggc tcccagaatt gttggtgagc aaattcaagt 540
tgctgcctgg gaagtcctga ctgacagagc agtaggtgac gctgacggag ctgagccgag 600
acttggggaa catctgaaac tncctgctca agctgagt 638

<210> 26
<211> 469
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(469)
<223> n = A,T,C or G

<400> 26
naggtacca atggagaaaa ctctttccgg agacgttcat catcaatacc atcatcaaga 60
tttttcacat aaagattaac accctgggat ctgggtgatcc tatcttggtt catctgttca 120
aatttgcgct taagttccgt ctgccgttcc acctttttct gagctcgacc aacataaatt 180
tgttttccat tgagctcctt tccgttcac tcatccacag ctttctgtgc atcttcatgc 240
ctttcaaagc ttacaaatcc aaatcctttg gattttccac ttcatcagt cattactttc 300
acacttaagg caggcccaaa cttgccaaag agatccttaa ggcgtcatc atccatgtct 360
tctccaaaat tcttgatgta aacattgggtg aattcttttg cctagctcca agttcagctt 420
ctcgtcttta cgagacttaa atcggccaac aaatactttg cgatcattt 469

<210> 27
<211> 364
<212> DNA
<213> Homo sapien

<400> 27
actctgctat ggtgctggct tcctttaaac tcaggataga tgccagggtgg gctccgtttc 60
cgtaagactg aactcgcagc tcggcatcag accagttcct cagcttcctg aagtaaccat 120
agcaattgga cttgtggtaa aaccatccag gagcacagct gggctctcatg atgatatac 180
ccaggactcc tgttttggcc aggcagctca gcaataggag cagccgcag cttctggaag 240
ccatcttcct cctaccctga ggatgtagct agtgcaagga tctcagagac cttactagcg 300
cttctttgaa actcctgggt tctccttgat ctgcaaatct gtttggcaac caagactcta 360
aggg 364

<210> 28
<211> 714
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(714)
<223> n = A,T,C or G

<400> 28
ccttcgagaa gatccctagt gagactttga accgtatcct gggcgaccca gaagccctga 60
gagacctgct gaacaaccac atcttgaagt cagctatgtg tgctgaagcc atcgttgagg 120
ggctgtctgt agagaccctg gagggcacga cactggaggt gggctgcagc ggggacatgc 180
tcactatcaa cgggaaggcg atcatctcca ataaagacat cctagccacc aacggggtga 240
tccactacat tgatgagcta etcatcccag actcagccaa gacactattt gaattggctg 300
cagagtctga tgtgtccaca gccattgacc ttttcagaca agccggcctc ggcaatcatc 360
tctctggaag tgagcgggtg accctcctgg ctcccctgaa ttctgtattc aaagatggaa 420
cccctccaat tgatgcccat acaaggaatt tgcttcggaa ccacataatt aaagaccagc 480
tggcctctaa gtatctgtac catggacaga ccctggaaac tctgggcggc aaaaaactga 540
gagtttttgt ttatcgtaat agcctctgca ttgagaacag ctgcatcgcg gccacgaca 600
agagggggag gtacgggacc ctgttcacga tggaccgggt gctgacccc ccaatggggg 660
actgtcattg gatgtcctga agggagacaa tcgcttttca tgctggtagc tggc 714

<210> 29
<211> 373
<212> DNA
<213> Homo sapien

<400> 29
acttgagatc cacagtcacg tgaactttgc cggctctctt acatctgccc acttcatttt 60
cattctttcc ttcccacaca atggtttttc caatgtgcaa gaatgatttc tcgacaaatt 120
cccggacact atggacctcc ccagtagcta taacgaaagc cttccgggtca tcattctgca 180

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| acatcaacca | catagcctcc | acatagtcct | tggcatggcc | ccaatctcgt | ttggcatcca | 240 |
| gatttcccaa | actgaaacat | tccagttgtc | caaggtaa | cttagctact | gaccggctaa | 300 |
| tttttcgagt | aacgaaatta | gcttctcttc | tgggactctc | atgattgaag | agaatgccgt | 360 |
| cactgcaaag | aga | | | | | 373 |

<210> 30
<211> 485
<212> DNA
<213> Homo sapien

| | | | | | | |
|------------|------------|------------|------------|-------------|------------|-----|
| <400> 30 | | | | | | |
| aaaactacga | ctcagcatac | atthttccac | atacattttt | acattgtacc | ttaggactca | 60 |
| gtcatctcca | cttaaattga | tgacacaagc | agctaataac | catttctggg | tttctgccta | 120 |
| acccccta | tgtctgttaa | agccaattct | ctgggtgtcc | cagtgaagg | tggtttttt | 180 |
| tctttccaca | ttggcacatt | cacttctccc | actcttggca | tgtaagaa | aagcatttac | 240 |
| ataattggaa | aaatctggat | ttctgatgcc | aaagggttaa | agcttcttgg | atttcatttc | 300 |
| attgatatac | agccactatt | ttatttttga | tcagtggcct | ttggggccact | gttcagggta | 360 |
| ctgaccatca | gtgtcagcat | tagggttttg | gtttttgttt | cttttgggtc | tttctttttt | 420 |
| ggcacatgtg | aatcttggtt | tgtgtaaa | gaaattactt | tctcttgttc | tctgatgatg | 480 |
| ggttt | | | | | | 485 |

<210> 31
<211> 342
<212> DNA
<213> Homo sapien

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| <400> 31 | | | | | | |
| acacattaag | catccccagt | tcccctcgca | caccctttt | cccagccact | agtaaccatc | 60 |
| cttctactct | ctatatccat | gagttcaatt | gttttgactt | ttagatccc | caaataattg | 120 |
| agaacatgca | atgtttgtct | gtttctggct | tatgtcactt | aatatagtga | cctctagttc | 180 |
| catccatgac | tccttaactg | cccctgaatt | tttgacacta | ttatttttaa | gtattttgga | 240 |
| aaactcacac | ctgttctcat | ttttaaacct | taataataac | aatttcctac | taagctaata | 300 |
| aaacttcccc | ttatattatt | tgtaatgtgt | gcataacata | gt | | 342 |

<210> 32
<211> 331
<212> DNA
<213> Homo sapien

| | | | | | | |
|-------------|------------|-------------|------------|------------|------------|-----|
| <400> 32 | | | | | | |
| acagtatgtg | gcatttccag | gtatgactga | gtgtgagaga | catgtcagag | gctcttcagt | 60 |
| gatttcttgc | tattgaccga | tgcttcaactg | tgccaaaaga | gaaaaaaat | gttgggtttt | 120 |
| gtaattaaat | tatttatata | tttttgaac | ccgaattgaa | aatgttgcag | gcaacgggct | 180 |
| acagctttat | tagtggttct | ctaactgtgg | tctccttggg | ccaagcaatt | tctttaaagg | 240 |
| aaaagttgat | tatgtatgtg | gagtgccagg | accactgcct | tgaaagcaag | tgtgattttt | 300 |
| atthtttaata | ttattttatt | tgtgtctgtg | t | | | 331 |

<210> 33
<211> 381
<212> DNA
<213> Homo sapien

| | | | | | | |
|------------|------------|-------------|-------------|------------|------------|-----|
| <400> 33 | | | | | | |
| acactgttgg | tgttatatgg | ggatgggggtt | ctcggttaatt | ttgtttatta | tttatgttta | 60 |
| ttattatgtt | ttatcattaa | ttattcaata | aatttttatt | taaaaagtca | ccctacttag | 120 |
| aatcttctg | tgggggtggg | agggacaaaa | gattacaaac | caaaactcag | gagatggtaa | 180 |
| cactggaatt | gataaaatca | cctgggatta | gttgataaac | tctgaaccac | caaacctctg | 240 |
| ttatcaagcc | ttgctacagt | catggctgtc | cagaaagatt | tacagttatt | tttctgagaa | 300 |

aggatccatg ggctttaaga acttcagaac tttaagaact tcagaagttc ttaagttgct 360
gaagctcaag taacgaagtt g 381

<210> 34
<211> 315
<212> DNA
<213> Homo sapien

<400> 34
acgaaactgt atgattaagc aacacaagac accttttgta tttaaaacct tgatttaaaa 60
tatcacccct tgaggctttt ttttagtaaa tccttattta tatatcagtt ataattattc 120
cactcaatat gtgatttttg tgaagttacc tcttacattt tccagtaat ttgtggagga 180
ctttgaataa tggaaatctat attggaatct gtatcagaaa gattctagct attattttct 240
ttaaagaatg ctgggtgttg catttctgga ccctccactt caatctgaga agacaatatg 300
tttctaaaaa ttggt 315

<210> 35
<211> 567
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(567)
<223> n = A,T,C or G

<400> 35
tacttcttaa aanacatata acacaatgtg gtagtagtag gtgtaaggaa ggtaagtttt 60
ttcatagtgg tatgcaaaca tatcattgaa atattacata gatataaaga cttagggaat 120
aaaaatagca gcaacaaata cttgatagat ttatcctact tgggagaaat attttgtagc 180
agagtattta gtatacttag aagttgattt agcaattagg ctttaatgac cttacaaagt 240
gaacataact gaacacaagt attttttcaa tgcaagatga ggatgaaaat ttacatttc 300
aaccatctg gctaaagtta agacttagca aaaattaaaa tgttgccttt gtccaagtat 360
agattaaggc aacaaacata tttgggtgtg taatttgaag ttttggactg aaatatcttt 420
gcaagtatcc acataaaatt ctgtaatgcc ttataattat attctaataa ttatgcatta 480
tactaagaca ccattaagaa cagttgange actacactaa atcaaaccat aaatgaggaa 540
aaaactttta atggtctttt ctagaag 567

<210> 36
<211> 265
<212> DNA
<213> Homo sapien

<400> 36
acaagtgggtg gccacagaag taggggggtc ttccttaagc tctgtgtcag agttccacct 60
gatccttatg gatgtgaatg acaaccctcc caggctagcc aaggactaca cgggcttggt 120
cttctgccat cccctcagtg cacctggaag tctcattttc gaggctactg atgatgatca 180
gcacttattt cggggtcccc attttacatt ttccctcggc agtggaagct tacaaaacga 240
ctgggaagtt tccaaaatca atggt 265

<210> 37
<211> 476
<212> DNA
<213> Homo sapien

<400> 37
actgtatgtg ttttgtaaat tctataaagg tatctgttag atattaaagg tgagaattag 60
ggcagggttaa tcaaaaatgg ggaaggggaa atggtaacca aaaagtaacc ccatggtaag 120

```

gtttatatga gtatatgtga atatagagct aggaaaaaaa gcccccccaa ataccttttt 180
aaccctctg attggctatt attactatat ttattattat ttattgaaac cttagggaag 240
attgaagatt catcccatat ttctatatat catgcttaaa aatcacgtca ttctttaaac 300
aaaaatactc aagatcattt atatttattt ggagagaaaa ctgtcctaata ttagaatttc 360
cctcaaattc gagggacttt taagaaatgc taacagattt ttctggagga aatttagaca 420
aaacaatgtc atttagtaga atatttcagt atttaagtgg aatttcagta tactgt 476

```

<210> 38
 <211> 424
 <212> DNA
 <213> Homo sapien

<220>
 <221> misc_feature
 <222> (1)...(424)
 <223> n = A,T,C or G

```

<400> 38
tacaagaacc tcaactcactg gacattgann ttctactgtc caatcccaac tnactgctgt 60
tnantggaaa cctgattctg gcagctcatt tatcttggtt tcttcatttg taaggctcgtt 120
cagttggact gatcatctct gagggccttg aagccctaac aagtctatca tgatcccaga 180
tgtaaaatat atatatgtgt atatataata tttcagctga gaagtgagtc ttcacaccaa 240
gtctactttt tgcaagttac tgggtttctg tcttcaccat cttctgaaaa gtctgcttct 300
gttggttcag tttctggggt catctgagta gagagattct gaaacagaca ctgatgttaa 360
tttgggggac tacttttctc atgcaaacag gggagctcct ancaatcctg agaggngctg 420
catc 424

```

<210> 39
 <211> 493
 <212> DNA
 <213> Homo sapien

```

<400> 39
acattgtagc cctctgctc tctaccctta acagctgcat cgacccttt gtctattact 60
ttgtttcaca tgatttcagg gatcatgcaa agaacgctct cctttgccga agtgtecgca 120
ctgtaaagca gatgcaagta tccctcacct caaagaaaca ctccaggaaa tccagctctt 180
actcttcaag ttcaaccact gtttaagacct cctattgagt tttccaggtc ctcagatggg 240
aattgcacag taggatgtgg aacctgttta atgttatgag gacgtgtctg ttatttccta 300
atcaaaaagg tctcaccaca taccatgtgg atgcagcacc tctcaggatt gctaggagct 360
cccctgtttg catgagaaaa gtagtcccc aaattaacat cagtgtctgt ttcagaatct 420
ctctactcag atgaccccag aaactgaacc aacagaaagc agacttttca gaagatgggtg 480
aagacagaaa ccc 493

```

<210> 40
 <211> 464
 <212> DNA
 <213> Homo sapien

<220>
 <221> misc_feature
 <222> (1)...(464)
 <223> n = A,T,C or G

```

<400> 40
acaaaacaca caaacatcac tttacttggg aaattatttt catcatactg taaacatctc 60
ttcccctaca tctggacatt ttgaaatagt ctttggtatt actagttatt gtgctttgaa 120
acagaaactt gcagaatttc tgtagtagtg ctacataaag atataaataa gaaaaatgca 180
cttggataaa gttacattta gctgcttttg cataattttc aaaaactaca gtgtatgcct 240

```

```

agtcacagtt ttatgagaaa gaatatttcc tttttcaact taattttaag gaacacttaa 300
tcattttggc taagtatcca tttttggagt ggatctgatg agttgcatga cactaaactt 360
ggatgctctc catttgctga aaggcacatt ttttaagaatg gattgnatag aagttgatcc 420
ttctggatct cccatatctg ctctccagtg acaactgnct tgtg 464

```

```

<210> 41
<211> 557
<212> DNA
<213> Homo sapien

```

```

<220>
<221> misc_feature
<222> (1)...(557)
<223> n = A,T,C or G

```

```

<400> 41
acagtgatag gtatctttct ttggagtttt ttttttgncc atatgtgtat agtttttatgg 60
gttctgagtt ggtgaccana aagttgcatg tagngctggc acttacttaa taactattca 120
tgatattgtt aataacttgt tataggattg tattcccaat tacagtctct aanattgtaa 180
ttgatattat ctganaggna gngngacaac tttcttttgt tgttacatta agccgaaaac 240
ataatactaa tagacaacta acagtttgct tatcaggcac atcaactaag gcacctcccc 300
ccatgctaag tttctcctgg atatatggaa gttgattgtt tcccagttna aaaacttgaa 360
ctaatatctc ctaaaaaaat ctgagtcctat attgttttta ttttacttag ctanaatctc 420
atagcangtt aaagtcatat ccttatcccc actaaaaata actatgtnta tgtgagagga 480
atatagtatg tgggagctgt attaaatact attacaggtg ttacagaatc tttaaataaa 540
tggacatgga ccaactt 557

```

```

<210> 42
<211> 255
<212> DNA
<213> Homo sapien

```

```

<400> 42
actatcaggc tttgtgctga tttcctgaac aaactgcatt atattatgaa aacaaaagga 60
aaagaagaaa taataaaaac tatactccca tatttcactt acagtgtttg agttcctgga 120
aggacctata taatggaggc agcattcaaa caagaaatta tgccaatcaa ctgtcaaatt 180
ttcactataa ttttcctaaa aaggcgtttt tcccccaata tctattaatc tcaaagaaac 240
ataagttgtg aatgc 255

```

```

<210> 43
<211> 349
<212> DNA
<213> Homo sapien

```

```

<400> 43
actccagcag atttaatat ggcatccatc atctagtcaa acctctcaca tgttcttcaa 60
atcaatcaaa tttgggattc tcaacatttt ctgtgtcaat aaaagggtgtg gaattagtag 120
attcgatgaa gacctgtttt tccttgccac attggacttc cagacgccat ttggattggg 180
tttagaagat ggggaaattt agaagacgtt tcttggcctg agtctcttaa gagtagagat 240
gcagaagaga gagtgagacc acgaagagac tggctgttga ctgcagggca ccaccagccg 300
ccttgggtgtt ggcattagtt ggatttgggg ccaaccaga gttggaagt 349

```

```

<210> 44
<211> 483
<212> DNA
<213> Homo sapien

```

```

<400> 44

```

| | | | | | | |
|------------|------------|------------|------------|------------|-------------|-----|
| accaaaccat | tttatgagtt | ttctgttagc | ttgcttttaa | aattattact | gtaagaaata | 60 |
| gttttataaa | aaattatatt | tttattcagt | aatttaattt | tgtaaagcc | aaatgaaaaa | 120 |
| cgttttttgc | tgctatggc | ttagcctgta | gacatgctgc | tagtatcaga | ggggcagtag | 180 |
| agcttggaca | gaaagaaaag | aaacttggtg | ttaggtaatt | gactatgcac | tagtatttca | 240 |
| gactttttta | ttttatata | atacattttt | tttccttctg | caatacattt | gaaaacttgt | 300 |
| ttgggagact | ctgcattttt | tattgtggtt | tttttggtat | tggtggttta | tacaagcatg | 360 |
| cgttgcactt | cttttttggg | agatgtgtgt | tggtgatggt | ctatgttttg | ttttgagtgt | 420 |
| agcctgactg | ttttataatt | tgggagttct | gcatttgatc | cgcacccctt | gtgggtttcta | 480 |
| agt | | | | | | 483 |

<210> 45

<211> 281

<212> DNA

<213> Homo sapien

<400> 45

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| acacgagaa | tccacgccc | gggaccagta | ggacttgagg | gactgcttac | tactaagtgg | 60 |
| ctgctgcgag | ggaaggacca | cgtggtctca | gatttctcag | agcatggaag | tttaaaatat | 120 |
| cttcatgaga | acctccctat | tcctcagaga | aacaccaact | gaaaagagcc | aggaaaaccc | 180 |
| gggaattttc | caaaaggtct | tcacgttaaa | cttgtcttat | ctcaggagag | agcccgtctt | 240 |
| tgtctcccag | ttcctggtag | ggtctgcctg | ttggaaagtg | t | | 281 |

<210> 46

<211> 587

<212> DNA

<213> Homo sapien

<400> 46

| | | | | | | |
|-------------|------------|------------|-------------|------------|------------|-----|
| acagcccggc | ctcccttgat | gcatttgggc | cgttccctgaa | aagttgtgtg | taaaggaaga | 60 |
| atttgccatc | aagccatttc | ccccttttgt | ttctaaaatt | atttcagaga | tgtgtgctcc | 120 |
| tggagggaaa | aagaaatacg | gcctcaacag | attaaaaaac | aaaagtcaca | cttaaggatc | 180 |
| cttctagtea | catcagcagt | gttctgcctt | tatgtagtag | ttgggcatat | aatccttcca | 240 |
| cacagcccct | gcagggaaag | gctaacttta | cggataatcc | acgtgagatt | tccacacaag | 300 |
| agaaaagcac | acgcatagt | aaatgtcagt | cttttcagta | atgaggatac | ctttaaggca | 360 |
| ctcttggaact | ctcggcaacc | acaacataat | agttgaaaga | tcaagattgg | ctccacgaaa | 420 |
| gtgatacgga | ggttaggatg | ctacttgctg | caaacaagcc | ctactttggc | caacatcctg | 480 |
| cttattttct | aaaaaagagg | gacagtgaag | acaaaaacga | cattgggaca | tgctgctcaa | 540 |
| ggtagttata | tatacgataa | gttgtatata | tgatcactgg | tagccta | | 587 |

<210> 47

<211> 317

<212> DNA

<213> Homo sapien

<400> 47

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| gaggactctg | acagccataa | caggagtgcc | acttcatggt | gcgaagtga | cactgtagtc | 60 |
| ttgtcgtttt | cccaaagaga | actccgtatg | ttctcttagg | ttgagtaacc | cactctgaat | 120 |
| tctggttaca | tgtgtttttt | tctccctcct | taaataaaga | gaggggttaa | acatgccctc | 180 |
| taaaagtagg | tggtttttga | gagaataaat | tcatcagata | acctcaagtc | acatgagaat | 240 |
| cttagtccat | ttacattgcc | ttggctagta | aaagccatct | atgtatatgt | cttacctcat | 300 |
| ctcctaaaag | gcagagt | | | | | 317 |

<210> 48

<211> 512

<212> DNA

<213> Homo sapien

<400> 48


```

acacttgat  ggcttttcac  cagtgtgagt  cctcaggtga  gcttttaa  gagaagactt  60
ggtataaact  tttgtgcaac  cagggtaatc  gcagtagtgg  atgcgtcg  tctccaaatc  120
ggggttactc  cttctattgt  atctgacagg  ttggatgttt  tgtgagttaa  ctggcaggg  180
ggtgggtaaa  tttggattgt  gaattgccag  tttagaagca  attgtagcag  cataggatgg  240
aggtgggggt  aaattctgga  gcatctctgc  ttgtctatct  ggacttccag  gctctgagct  300
tggtgggtgac  gggggaaagt  aagtggcctg  ttgtggaaga  aactgacttg  gcattgtgta  360
tgtgcaagg  ggcattgcc  ggaattgttt  cactgcagtc  tgcggaacag  cagaggtgtg  420
tgtgttaagg  cctgccatgg  cagctgacat  agaaacatta  agagtgtcca  ttgctgctgt  480
ctgatttgta  gaactgggca  tatctagatc  cg  512

```

<210> 49

<211> 454

<212> DNA

<213> Homo sapien

<220>

<221> misc_feature

<222> (1)...(454)

<223> n = A, T, C or G

<400> 49

```

acaggattca  ctaactgttt  cgaatgaagc  ccaaactgcc  aaggagttaa  ttaaaatcat  60
agagaatgca  gaaaatgagt  atcagacagc  aattagttaa  aactatcaaa  caatgtcaga  120
taccacattc  aaggccttgc  gccggcagct  tccagttacc  cgcacaaaaa  tgcactggaa  180
caagatactc  agctacaaga  ttggcaaaga  aatgcagaat  gcttaaaggc  tgaatgtagg  240
attcttcagt  atgtggaaag  acaaggattc  aacgtgtggt  catatgataa  ataagtgttt  300
tataaacaag  agtgatattt  tgctagggtt  ttcaaagtta  accggttttc  tagcctcatg  360
gaatactgtt  gaacctatag  cggtgtcttg  attcttttgt  gttctctgcc  ttgtaatttt  420
ctgttactgc  tatactctac  tgtaaattct  tntt  454

```

<210> 50

<211> 374

<212> DNA

<213> Homo sapien

<400> 50

```

actatcccat  gttgcgcagt  aatagatggc  ctcgtcccca  gtccggagtc  cgggtgatggc  60
cagggcggt  gacgtgccag  acttggtggc  agagaatcgg  tcaggaattt  ctgagggacg  120
gccatcattg  tgataaatga  ggagtttggg  ggctgttcct  gagaattgta  gataccacga  180
cacataatta  gttccaatgt  tggaggcgt  tccagagcag  gacatggaga  ccttctgtcc  240
tggggcgcga  gagactgagg  gcggctgcgt  caagatggac  tgggcccagg  accctgtgca  300
gtgaatgaga  aggggtgagga  ggagagggga  gcaggtcatg  atgaagattg  tcccagatcc  360
tgccttctgc  gctc  374

```

<210> 51

<211> 250

<212> DNA

<213> Homo sapien

<400> 51

```

accagatatt  ttctatactg  caggatttct  gatgacattg  aaagacttta  aacagcctta  60
gtaaattatc  tttctaattg  tctgtgagga  caaacattta  tggttcagatt  gaaattttaa  120
ttaatatcat  tcaaaaggaa  acaaaaaatg  ttgagtttta  aaaatcagga  ttgacttttt  180
tctccaaaac  catacattta  tgggcaaatt  gtgttcttta  tcaattccga  gcaaatactc  240
agatttataa  250

```

<210> 52

<211> 351

<212> DNA

<213> Homo sapien

<400> 52

| | | | | | | |
|------------|-------------|------------|------------|------------|------------|-----|
| acgaaagggt | ttgtaccaat | attcactacg | tattatgcag | tatttatatc | ttttgtatgt | 60 |
| aaaactttta | ctgatttctg | tcattcatca | atgagtagaa | gtaaatacat | tatagttgat | 120 |
| tttgctaaat | cttaatttaa | aagcctcatt | ttcctagaaa | tctaattatt | cagttattca | 180 |
| tgacaatatt | tttttaaaaag | taagaaattc | tgagttgtct | tcttggagct | gtaggtcttg | 240 |
| aagcagcaac | gtcttttcagg | ggttggagac | agaaacccat | tctccaatct | cagtagtttt | 300 |
| ttcgaaaggc | tgtgatcatt | tattgatcgt | gatatgactt | gttactaggg | t | 351 |

<210> 53

<211> 546

<212> DNA

<213> Homo sapien

<400> 53

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| acatggacat | tctgcaaacc | cagctgtcac | atTTTTcttg | caactccttt | tgcaaaagca | 60 |
| gactaaaatg | ttttaaaatg | tgaaaaaaca | ttatTTTTtc | aaagcaagaa | aataatttac | 120 |
| tgccctctta | cataatgtat | ttataaagtt | tttccagata | aactaatcaa | ataaattaga | 180 |
| ataatgtgac | aacattacaa | atttaatttg | ttagctgcat | tccttctgat | gttaccacga | 240 |
| tagaatgtta | ctgatgattc | agggtctatt | ctgaagtctg | tatgttgctg | ctgtccccag | 300 |
| tgatggtgga | cttatctttg | ccttacctga | tcacaaatta | tggtggggaa | aataaagatt | 360 |
| taatatttct | ttaaatagaa | aaagaatttg | gttttgctcg | tttaagagea | atgagaaaat | 420 |
| gatggaatgt | tgactgtgtt | tggcacacag | gacacggacc | ttcatggag | tccttgctct | 480 |
| gcgtggcatc | tgtcagcttt | tcacctttca | ttcttattct | tcaettttgc | tgctgagcct | 540 |
| agctgt | | | | | | 546 |

<210> 54

<211> 631

<212> DNA

<213> Homo sapien

<220>

<221> misc_feature

<222> (1)...(631)

<223> n = A,T,C or G

<400> 54

| | | | | | | |
|------------|------------|-------------|-------------|------------|------------|-----|
| acngttttta | ccaatacnna | naagcantaa | agcaataata | tctgaagcat | tatttaagaa | 60 |
| atctcaatac | acgatctctg | aagttcctaa | aattctggca | ctaattctaa | tgtgaactta | 120 |
| gtagcaaaag | accagaaat | agtaagccct | tgacctaaaa | actaactgat | ttgtatgata | 180 |
| ttcatgcaga | aacaatgatg | aaatggagtc | aagttttcta | gtgtcattgt | tatcaaaata | 240 |
| actgtcaaaa | tagtaagttt | gaaacttaaa | tgagcacaaa | ataaaatttt | gttttctaac | 300 |
| aagaccagat | ttctttttta | aaataattct | gagttagaca | aagtgatttt | cctaaaagct | 360 |
| agctgaagct | accttaaata | tccctatttt | taagttacag | catctctaaa | taagttaatc | 420 |
| acacaagata | gtttaaatac | acctttaggt | gtagggggagg | ggagaagcgc | ctctttttct | 480 |
| aatgcagctg | ttttaatttg | aagcttttgc | acaaaatcag | atagaaacat | taatgcctaa | 540 |
| ctcataatga | cccttgatta | cttghtaattt | tggactagaa | ataatgtggc | tttgaacatg | 600 |
| ccagtgttag | accatactga | cttaaaaaaa | t | | | 631 |

<210> 55

<211> 408

<212> DNA

<213> Homo sapien

<400> 55

| | | | | | | |
|------------|------------|------------|------------|------------|------------|----|
| accaatatat | ccccagaaag | aattgcaatt | taccaagggt | ttcacgtgtt | ttgagagaaa | 60 |
|------------|------------|------------|------------|------------|------------|----|

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| tcttactgaa | agactagtga | tgtccatfff | ccagtaaata | ctgagcgaaa | aacaattfff | 120 |
| ataccccaat | ctgaggata | aacttgctff | ttgtgggata | acaactgctg | taaattagac | 180 |
| aattgtagca | acaatccaag | acaataacag | aatgcctatg | acagtctgcc | atattctggg | 240 |
| gagtgtctat | caaagctcat | catgattfff | tgtgagatct | tccccgtaat | tggtagcttg | 300 |
| gcttccaaca | aacatgttcc | agttctccaa | tatttctctt | ttagttagct | tctcatcctt | 360 |
| gtttttgtct | gattcatata | ccagatgcct | ggcctcagcc | tgtgcgtg | | 408 |

<210> 56

<211> 567

<212> DNA

<213> Homo sapien

<400> 56

| | | | | | | |
|------------|------------|------------|------------|-------------|-------------|-----|
| actgtgggct | gaagtaatgg | atacggacgt | aaccatcttc | gccgccgctg | ctgtagctct | 60 |
| tgccatcagg | atggaaggca | acactgttga | taggtccaaa | gtgacccttg | actcttccaa | 120 |
| actcttcttc | aaaggccaaa | tggaagaacc | tggcctcaaa | cttgccaatc | ctgggtggagg | 180 |
| ttgtgggtac | atccatggct | tcctgaccac | cgcccaggac | cacatgggtca | tagttggggg | 240 |
| agagggcagc | tgagttgaca | ggacgttctg | tccggaaagt | cttctgatgt | tcaagagttg | 300 |
| tgaggtcaaa | aagcttggct | gtgttgcctt | tggacgcggt | cacaaacatg | gtcatgtccc | 360 |
| tgataactg | gatgtcgttg | atctgccggg | agtgtcctt | aacattcacc | aacacctctc | 420 |
| cagacttggc | actatactgg | ttgagctctc | cactctcatg | gccagcgatg | atgcactccc | 480 |
| ccaggggtcc | ccaaacagca | ctgggtgatt | tagagtcatt | gcaagggatc | ttcatgtagg | 540 |
| gctcattggg | gtcaatctgg | ctcggat | | | | 567 |

<210> 57

<211> 411

<212> DNA

<213> Homo sapien

<400> 57

| | | | | | | |
|------------|------------|------------|------------|-------------|-------------|-----|
| acccttcctt | gtccgaagga | gctgaccagt | attgatgaga | gagtcacaggc | agctcctgaa | 60 |
| gttcagctgg | tagtttggtc | tctgaacatt | tggtctcttg | aaggcacagt | atafctgggg | 120 |
| cttcttcctt | taccaatct | aatcctttct | tcttaatcca | ggctcgaagc | ccatccacat | 180 |
| tccaagagca | gatcttgagt | gtggcagggt | tgccactggg | tgaggttttc | tgatctgggg | 240 |
| ggtcctcata | cagggtctgg | ccctctcctg | ctgcctcttt | gtcatttttc | tttgccggccg | 300 |
| tcttactctt | cttggcctct | ggctctgtcc | tgagctcatc | cccgtcttcc | gccaccgctc | 360 |
| cctttttccc | acgtctcgge | attcccgtta | cgaacgcctt | tgggcagctg | t | 411 |

<210> 58

<211> 589

<212> DNA

<213> Homo sapien

<220>

<221> misc_feature

<222> (1)... (589)

<223> n = A, T, C or G

<400> 58

| | | | | | | |
|------------|------------|------------|------------|-------------|------------|-----|
| acattaatac | aaacatactt | gcagtctgag | cgaagatggg | aatggaggct | gaggaggtca | 60 |
| aaggacgaaa | ggtcagccct | aaagacaggg | tgttttggtt | ttatggtaat | tacaccttca | 120 |
| taccttctat | aatattcatt | gacagacggg | gacatcaaca | gggtgtagttt | atcatgttct | 180 |
| gtgtagagaa | ctaaactacc | ctactgtatt | tgccatgccc | ccaattccaa | gaaaacggca | 240 |
| aaaaattagc | ccatcccatt | cctcatcaca | aagatcttaa | ctgcacccct | gcaacacaag | 300 |
| acttttccaa | taggacaaaa | cttcaaacag | cattgtatac | caaatgattg | cggatcaaaa | 360 |
| ttaaatttac | aggaacacaa | tactgaagca | ctccactgtt | gctgtaaaaa | ctgctggaaa | 420 |
| cagaatctgt | caactggcca | aattttatcc | ttaattatta | tccaaacagc | cgctcctctc | 480 |
| acatctatcc | ggatgatgct | aatctactac | cctgtccact | aggttagcaa | gttgtaggaa | 540 |

caactcttca ccatttctcc caccctaaga ggtacctgcc cnggcggnc 589

<210> 59

<211> 440

<212> DNA

<213> Homo sapien

<400> 59

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| acatgaggca | ggtgagcagc | actggagaac | cttcacggtc | cacacggaac | tccccagttg | 60 |
| gagtataata | gtcattctcc | ttgatatggt | tgcctgtatc | tgtgctccct | ccaatccgga | 120 |
| ccatccaaag | aaacttggtg | atatcatcag | aggaataccc | agtgaggcct | ccaaaaatga | 180 |
| ccagcacata | gctgacatcg | agctccctca | tgatctcata | ggctttttcc | tctgtggacg | 240 |
| ccattgcctg | ccctactcga | gaaatatggg | tattattcca | tgtgttattg | tccactaaaa | 300 |
| ttgttcgggt | tgccatagct | gtaatctgat | agccataatc | ccaccaggac | atgaccttcg | 360 |
| catcctctgg | agtattatga | cgaagccaat | aatatgcttc | tcggaagtca | tcaaatatga | 420 |
| tcctactgcc | atccccacca | | | | | 440 |

<210> 60

<211> 417

<212> DNA

<213> Homo sapien

<400> 60

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| acctggaaga | tcaagatcta | cagctgecta | tttccacatc | tttcaatcca | tctggctcct | 60 |
| taaatagggg | aaaaagccct | tatttggtgg | agaagcattt | ccaaaatgaa | gttacagggt | 120 |
| ctattaaaac | ttactgtcac | atcaactggt | aaaatagggc | cttttgtggt | ttgttatttc | 180 |
| accttaatat | caccagaatt | cctgtaattc | cacaattgtg | attttactat | gtagaagata | 240 |
| attcagttct | agtctattgc | tttagatgta | aaaacagctg | aaaacccaaa | gtggattaga | 300 |
| attgctgaag | gatttccctg | ccgttggttg | atacaatcta | ttctcttgat | tcttgatagg | 360 |
| tgcatagaaa | gcctaactta | aaattctttc | tacaggaaca | tgtctgattt | caggagt | 417 |

<210> 61

<211> 354

<212> DNA

<213> Homo sapien

<400> 61

| | | | | | | |
|------------|------------|------------|------------|------------|-------------|-----|
| acctcctgtg | ttgcagagtt | tctttatcca | catccaccca | accagcagca | tcagccacag | 60 |
| gactggctct | gaggacatct | ggtgggctca | ttggaggtgt | gacatgaagg | atttcatatg | 120 |
| aatcacttg | ggtctctcct | ggtttgtcca | ggttctcaaa | tacagcctct | tgtttatcgg | 180 |
| ctcggacttc | aatgagggtt | ttcttgtagt | taacagttag | gttccgctcc | tggtatgatct | 240 |
| cctgcagggc | atctgcatac | ttcttaacct | cgaaaatggc | tccaagagaa | gtgttgaaaa | 300 |
| tgatattggc | cttggaatgc | ttccctgtct | tcctgaagta | ggcttctgat | aagt | 354 |

<210> 62

<211> 205

<212> DNA

<213> Homo sapien

<400> 62

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| accccccttc | acttcgtctc | ccctagctcc | tagaagcaac | cactgatgtg | atttctacca | 60 |
| aatccagttt | tggtcctact | aaatatactc | ttttgagact | ggcctctttt | actcaccata | 120 |
| atgcctttgt | aattcatcca | tgctgttggt | tgatcagca | gtttgttcct | tttcattgct | 180 |
| gagtagtatt | ctattgtaga | gatgt | | | | 205 |

<210> 63

<211> 325

<212> DNA

<213> Homo sapien

<400> 63

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| acacacgggt | tccggatcaa | tgctcggggc | aacgccactg | cctgtcgtg | acccctgac | 60 |
| agctggctcc | cagcctcgtc | tacctctgtg | tcatagccct | gagggagtcc | agagatgaaa | 120 |
| ctatggggcc | cagactttac | tgacagcagc | gtgatttcct | ccatagttgg | cttctgggtc | 180 |
| aggccatagg | caatattttc | ttgaagactt | cttccaaata | cctgtggctc | ttgtcccact | 240 |
| gcagccacct | gcctgtgcag | gtagcgggtg | tcatattggg | gaaggggctt | cccatccaac | 300 |
| agcagctgtc | ccccggtggg | ctgggt | | | | 325 |

<210> 64

<211> 599

<212> DNA

<213> Homo sapien

<220>

<221> misc_feature

<222> (1)...(599)

<223> n = A,T,C or G

<400> 64

| | | | | | | |
|-------------|------------|------------|------------|------------|------------|-----|
| actttgatgt | ttgaacaacc | ttttcttgat | cacttcttcg | caataaaaat | atgacatatg | 60 |
| tagtaaacct | taaaaaattt | cgtgtaactt | tatggctcta | cgctggaatt | cttctgaagt | 120 |
| gagtaatcat | cacaatcatc | tttagtatat | aatggatcaa | aatgacacga | ttgcaaatat | 180 |
| tgataacaca | cagttataaa | aggtgaaatt | ctattgggaa | cacatctctt | agtgagatag | 240 |
| atggggctga | cccaccaatt | aattcattta | tctggatgaa | tagttcctac | tggtagatta | 300 |
| acagggttca | ttttcaattc | tgttgttttc | acagatacaa | gtgctgagaa | atgggtttac | 360 |
| ataaatagggt | gagaatgcta | gtagttttgt | tgtaagcatg | tcaatcaatc | gtttggtttc | 420 |
| tttccgagtt | gcatgccaaa | aaccaaatag | tgttccttca | tcagctgaca | attcatgggc | 480 |
| caccattaat | tttgttgaaa | gcaaagaact | ggaaaccatc | tgacttgaaa | agaatttggt | 540 |
| atcctgggtat | tagaggcatt | cactttctct | agngactttt | aattatacta | attactctc | 599 |

<210> 65

<211> 373

<212> DNA

<213> Homo sapien

<400> 65

| | | | | | | |
|------------|------------|-------------|------------|------------|------------|-----|
| acattaaagt | gtgatacttg | gttttgaaaa | cattcaaaca | gtctctgtgg | aaatctgaga | 60 |
| gaaattggcg | gagagctgcc | gtgggtgcatt | cctcctgtag | tgcttcaagc | taatgcttca | 120 |
| tcctctctaa | taacttttga | tagacagggg | ctagtcgcac | agacctctgg | gaagccctgg | 180 |
| aaaacgctga | tgcttggttg | aagatctcaa | gcgcagagtc | tgcaagttca | tcccctcttt | 240 |
| cctgaggtct | gttggttgga | ggctgcagaa | cattggtgat | gacatggacc | acgccatttg | 300 |
| tggccatgat | gtcaggctcg | gcaacaggct | ccttggtgac | actcaccaca | ttgtttttca | 360 |
| agctgacttc | cag | | | | | 373 |

<210> 66

<211> 520

<212> DNA

<213> Homo sapien

<400> 66

| | | | | | | |
|------------|------------|-------------|------------|------------|------------|-----|
| acgtgagcca | gtcatccata | cactaaggcc | tagttgagaa | aaacctttga | ttcaggatgg | 60 |
| ctgggttact | aaccttgaaa | tgtaagagat | ctgggttttg | atgtaaaagt | tgcaacacac | 120 |
| aaacggaagt | cttaaaaact | ttttgctctg | gtcagttaca | ggtggatccc | caataatctg | 180 |
| tttttggttt | tctgatggaa | ataatagaat | taggggaaat | caaactctgt | tggtaggtgt | 240 |
| ctacagtatt | agaagagggt | ataagggcac | tgtttaacac | taagttctaa | tacttccaga | 300 |
| aactgtgcat | tccagatcta | cataactaaat | gctcttatca | ttttgaaatg | ggctcttgat | 360 |

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| taatagaccc | atatttttta | gtggcttcta | tgttgatat | ttgtctaaaa | tgaaagctct | 420 |
| tttgcgttct | aaaactacaa | tatatgtcat | cttattttcc | ctgagtatcc | aagtatagtg | 480 |
| cagattctat | gtaaaactac | taaatgacac | tggaatatgt | | | 520 |

<210> 67
<211> 241
<212> DNA
<213> Homo sapien

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| <400> 67 | | | | | | |
| acagagatgg | agaacgaatt | tgtcctcatc | aagaaggatg | tggatgaagc | ttacatgaac | 60 |
| aaggtagagc | tggagtctcg | cctggaaggg | ctgaccgacg | agatcaactt | cctcaggcag | 120 |
| ctgtatgaag | aggagatccg | ggagctgcag | tcccagatct | cggacacatc | tgtggtgctg | 180 |
| tccatggaca | acagccgctc | cctggacatg | gacagcatca | ttgctgaggt | caaggcacag | 240 |
| t | | | | | | 241 |

<210> 68
<211> 487
<212> DNA
<213> Homo sapien

| | | | | | | |
|------------|------------|-------------|------------|------------|------------|-----|
| <400> 68 | | | | | | |
| actttgaggg | attggtgggc | ttgggcccct | cctggcccag | gagatgtaga | atacgggtgg | 60 |
| ccagcactgt | gaactcgcag | tcctcgatga | actcgcacag | atgtgacagc | cctgtctcct | 120 |
| tgtctcttga | gttctcttca | atgatgctga | tgatgcagtc | cacgatagcg | cgcttatact | 180 |
| caaagccacc | ctcttcccgc | agcatgggtga | acaggaagtt | cataaggacg | gcgtggttgc | 240 |
| gaggatattt | ctgacacagc | gcactgatgg | cctggacaac | caccaacttg | aattcatccg | 300 |
| agatttctga | catgaaggag | gagatctgct | tcatgaggcg | gtcgatgctg | ctctcgctgc | 360 |
| ccgtcttaag | gagggtggtg | atggccagcg | tggcaatgct | gcggtttgaa | tctgtgacca | 420 |
| ggttctccag | atccagatta | caagctgtca | cagctgacgg | atgcttcatg | gcaaccttat | 480 |
| tgagggt | | | | | | 487 |

<210> 69
<211> 415
<212> DNA
<213> Homo sapien

| | | | | | | |
|------------|------------|------------|-------------|-------------|------------|-----|
| <400> 69 | | | | | | |
| actagcttca | agaagctttt | ggtcagctac | atttaaaggc | acaatagggc | ctttggattc | 60 |
| tttgtgtgta | attggttttt | cactgagtgg | tttggaaagta | tctaaatcgg | actttttact | 120 |
| atattccaca | cttactacca | catccttggc | gccaggagat | ttctcttggc | atgacaataa | 180 |
| ttcttcttgt | ccttgaagat | gagatatatc | cagaccttct | tttaggcgaa | taaccactac | 240 |
| tccatattgt | atgtcaaaag | catcatgaaa | taagtttata | tacatatcca | catccctcat | 300 |
| atctgcttgc | aaccaatctt | tcttaaattc | aaggacaagt | gtgtttggct | tcatacgacc | 360 |
| aagaccagca | gcctgcatca | aatactgtgc | accttctctc | aagtcattctg | catgt | 415 |

<210> 70
<211> 535
<212> DNA
<213> Homo sapien

| | | | | | | |
|------------|------------|------------|-------------|------------|------------|-----|
| <400> 70 | | | | | | |
| acatcatgtc | ttataaggaa | gccattaagg | tcaactccact | gccatgtatg | caactgctgt | 60 |
| gtggctcgat | atgatcaaca | ctgcctgtgg | actggacggc | gcatagggtt | tggcaaccat | 120 |
| cactattaca | tattcttctt | gtttttcctt | tccatgggat | gtggctggat | tatatatgga | 180 |
| tctttcatct | atttgtccag | tcattgtgcc | acaacattca | aagaagatgg | attatggact | 240 |
| tacctcaatc | agattgtggc | ctgttcccct | tgggttttat | atatcttgat | gctagcaact | 300 |
| ttccatttct | catggtcaac | atttttatta | ttaaataaac | tctttcagat | tgcccttctg | 360 |

ggcctgacct cccatgagag aatcagcctg cagaagcaga gcaagcatat gaaacagacg 420
ttgtccctca ggaagacacc atacaatctt ggattcatgc agaacctggc agatttcttt 480
cagtgtggct gctttggctt ggtgaagccc tgtgtggtag attggacatc acagt 535

```
<210> 71
<211> 249
<212> DNA
<213> Homo sapien
```

| | | | | | | |
|------------|------------|------------|------------|------------|-------------|-----|
| <400> 71 | | | | | | |
| agcgggacga | ggatgacgag | gcctacggga | agccagtcaa | atacgacccc | tccttttcgag | 60 |
| gccccatcaa | gaacagaagc | tgcacagatg | tcctctgctg | cgtcctcttc | ctgctcttca | 120 |
| ttctaggtta | catcgtggtg | gggattgtgg | cctggttgta | tggagacccc | cggcaagtcc | 180 |
| tctaccccag | gaactctact | ggggcctact | gtggcatggg | ggagaacaaa | gataagccgt | 240 |
| atctcctgt | | | | | | 249 |

```
<210> 72
<211> 297
<212> DNA
<213> Homo sapien
```

| | | | | | | |
|------------|------------|------------|-------------|------------|------------|-----|
| <400> 72 | | | | | | |
| acacactgat | tgtgcggcc | gacaacacct | atgaggtgaa | gattgacaac | agccaggtgg | 60 |
| agtccggctc | cttggaagac | gattgggact | tcctgccacc | caagaagata | aaggatcctg | 120 |
| atgcttcaaa | accggaagac | tgggatgagc | gggccaaagat | cgatgatccc | acagactcca | 180 |
| agcctgagga | ctgggacaag | cccgagcata | tccccgaccc | tgatgctaag | aagcccgagg | 240 |
| actgggatga | agagatggac | ggagagtggg | aacccccagt | gattcagaac | cctgagt | 297 |

```
<210> 73
<211> 531
<212> DNA
<213> Homo sapien
```

| | | | | | | |
|-----------------------|------------|------------|-------------|-------------|-------------|-----|
| <400> 73 | | | | | | |
| acttgtccca | ctcctgttca | gaggtcacat | gcttatccaa | aaactctgcc | atcccaatgc | 60 |
| ccattctccg | gcaaagtgcg | gcaatcactg | tttgggtattt | ctcagccaga | tttctaaact | 120 |
| caagggagat | cgttgggaag | tcctccagca | cctggcgatc | cttctccttg | ctctccatga | 180 |
| accgccagtc | tggttggtaa | aggaaagagt | gaaagttgtg | taacagcggg | accttctttt | 240 |
| ccacactgat | ggtcatgtca | tcttccagtg | tgtccagagc | tcggaagaacc | agataaaata | 300 |
| tgcacactgc | gttgcgcat | tccccatcca | gcgcctggat | aacagctgcg | aaactgcgac | 360 |
| tgggtctgatt | gagatacttg | tagcaagttt | tcaggctgct | gctgagcgag | tcctgggtcca | 420 |
| tcttgggcat | caccttccgc | ttgcccccca | tccggaagcg | caccaggttg | tagaactctt | 480 |
| cgggggtggcc | aaggcatttc | acgaactcca | tcctgggtgca | ggcggcggac | t | 531 |

```
<210> 74
<211> 394
<212> DNA
<213> Homo sapien
```

| <400> 74 | | | | | | |
|-------------|-------------|------------|------------|-------------|-------------|-----|
| actaaaactt | acaataaata | tcagagaagc | cgttagtttt | tacagcatcg | tctgcttaaa | 60 |
| agctaagttg | accagggtgca | taattttcca | tcagtctgtc | cttgtagtag | gcaggggcaat | 120 |
| ttctgttttc | atgatcgga | tactcaaata | tatccaaaca | tcttttttaa | actttgattt | 180 |
| atagctccta | gaaagttatg | ttttttaata | gtcactctac | tctaatacagg | cctagctttg | 240 |
| ctcatttttg | agcctcacta | aaataacaga | tttcagtata | gccaaagttca | tcagaaagac | 300 |
| tcaaattggaa | tgattttacaa | aatagaacac | tttaaaccag | gtcagtccta | tctttttgta | 360 |
| gctgaaggct | atcagtcata | acacaatttc | gcgt | | | 394 |

<210> 75
<211> 369
<212> DNA
<213> Homo sapien

<400> 75
acattggtga tcggagtata gttggagcgc tttgtcatga tttccagggt ggctttgtcc 60
acagctatgt tggccaatgc acctgagcc tcaaagctgg caaatcgccc aaattcttca 120
agccgccaga ccgtctcctt ctttgccata tccacatgga aaatctcatc accatcaaag 180
tcaaacataa actcgccctga ttggtcagga ttcagataga actcggcctg gatgatcaca 240
tgttcttctt tgatagccca tgattcctga gcgctcatca gcacagctat gatgaaaaat 300
cctagcacag ggactccact tatggccatt ttcttcttgg gcgctctgtt gggagtcagt 360
agagctcgg 369

<210> 76
<211> 384
<212> DNA
<213> Homo sapien

<400> 76
acgactcggg gctcgccctg tccgcggcct tgcaggccac tcgagcccta atggtggtct 60
ccctgggtgct gggcttctct gccatgtttg tggccacgat gggcatgaag tgcacgcgct 120
gtggggggaga cgacaaagtg aagaaggccc gtatagccat ggggtggaggc ataattttca 180
tcgtggcagg tcttgccgcc ttggtagctt gctcctggta tggccatcag attgtcacag 240
acttttataa ccctttgatc cctaccaaca ttaagtatga gtttggcctt gccatcttta 300
ttggctgggc agggctctgcc ctagtcatcc tgggagggtgc actgctctcc tgttctctgc 360
ctgggaatga gagcaaggct ggggt 384

<210> 77
<211> 291
<212> DNA
<213> Homo sapien

<400> 77
acgtggcagc catggctccc ttcacaagct gtaggctcctg gtgggacagc tggctttggg 60
gaagcttgct tttctgggtg acctatggat gctgcagaac ctgcttagct gtgaggcgct 120
ggtggggatc cacgtgtagc atcttggaca ccaggctcctt ggctgtctct gaaactgtgt 180
tccaatttcc cccactgagg gtaaaacttc cactgccgat ccgggttagg atttctctct 240
gtgtgtcact gggaccgttg gcaaattggag tatatcctgc cagcatggtg t 291

<210> 78
<211> 242
<212> DNA
<213> Homo sapien

<400> 78
acctatattg ctaatgctag gatcaagata ccacatagcc agaacaagaa gttgaaggta 60
aacatagaat attttataca ggcactcaca cctgccattt cggaaaagga ttaggaatcc 120
agatgccgtg aatttaacta ttcgttacag gcttgtcctg caatatgctc tggagcaact 180
tgctgcaga gatttctgta tccacggaca tttaaataac gcaaaggcta tctccaggca 240
ag 242

<210> 79
<211> 449
<212> DNA
<213> Homo sapien

<220>

<221> misc_feature
<222> (1)...(449)
<223> n = A,T,C or G

<400> 79
ngtacagaca aaactacaga cttagtctgg tggactggac taattacttg aagganttag 60
atagagnatt tgcactgctn aanagtcact atgagcaaaa taaaacaaat aagactcaaa 120
ctgctcaaag tgacgggttc ttggttgtct ctgctgagca cgctgtgtca atggagatgg 180
cctctgctga ctcagatgaa gacccaaggc ataagggttg gaaaacacct catttgacct 240
tgccagctga ctttcaaacc ctgcatttga accgaccaac attaagtcca gagagtaaac 300
ttgaatggaa taacgacatt ccagaagtta atcatttgaa ttctgaacac tggagaaaaa 360
ccgaaaaatg gacggggcat gaagagacta atcatctgga aaccgatttc agnggcgatg 420
gcatgacaga gctagagctc ggnccagcc 449

<210> 80
<211> 490
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(490)
<223> n = A,T,C or G

<400> 80
acatttcctt gnagactctg ntaatttcct gcagctcctg gttggttctg gagcagatga 60
tctcaatgag agagtcctcg tgggttccca gcccttcat ggaagctttt agctcanaag 120
cgtcatactg agcaggtgtc ttcaataggc ccaaaatcac cgtctccagg tggccagata 180
aggctgactt cagtgtctgat gcaagttcct ttttggtcct tctctggtag gcgaaggcaa 240
tatcctgtct ctgtgcattg ctgcgntgg tcaaaatggt gacaatggtg acctcatcca 300
cacctttggt cttgatggct gtttcaatgt tcaaagcatc ccgctcagca tcaaagntag 360
tataggcttt gacagaccca tatgcacttg ggggtgtaga gtgatcacc tccaagctga 420
gcttgacacag gatttcgtga acagtagaca ttttgaagga agctgggccg tgcgccgaga 480
gctgagagcg 490

<210> 81
<211> 339
<212> DNA
<213> Homo sapien

<400> 81
acagtagtaa ctgatgtccc cttcttcctg gatgaatgag cagataaata ttgatgtcag 60
catccttgaa ccatatcaaa gtgagcagtg tttggctact gcttctattt gaaatgggtg 120
tgtgttttgg ttgtggtctg aagctttgaa gcgctactta gcatctcctt tcttccatgg 180
agctctcacg attcaaacaat gacagatttg gtaaaatgct gggttaggttg agtcttcctt 240
gccccactc agtcattctt gtatgaatcc catgatttgg gggttttttt cttttttttt 300
ataccagttt ttagctggtg tttatgaaga acagtgagt 339

<210> 82
<211> 239
<212> DNA
<213> Homo sapien

<400> 82
caagaacagc taaaatgaaa gccatcattc atcttactct tcttgctctc ctttctgtaa 60
acacagccac caaccaaggc aactcagctg atgctgtagc aaccacagaa actgcgacta 120
gtggtcctac agtagctgca gctgatacca ctgaaactaa tttccctgaa actgctagca 180
ccacagcaaa tacaccttct ttcccaacag ctacttcacc tgctcccccc ataattagt 239

<210> 83
<211> 528
<212> DNA
<213> Homo sapien

<400> 83
acattcggtta ttttaaataa acaagttttac aaagttttatt ttcattctata cgtaaggatg 60
atTTTTTTTaa aactTTTTTtac atatttagtgg ttatgatcca atgtgtcatg agtgaattta 120
actgtaagggt ggttttaaatac aaatatgcaa tgtttacttg aattgtattt ctatttagcag 180
atTTTtgacta tgtttacagg acggttttaa ttaaggatta tcaggcatgt gagatctttc 240
agttatcttt aaagtagatg tatattaagg gcttagattt aggatctaca tattctgggc 300
attgaatagg cagtaactta caaataagtt ttgcttacct tttgttctag ggactagcac 360
tgctatcaat ggaaagtatt tttaactaat ctgttattaa gaaagtcata tttttgcatt 420
tcagccaaaa taaagaccgc ctgtaataat ctgttagaaa cagataatac atgtctgaaa 480
tccatatggt tcatatgatc taaactgtat tttccaattt aaattaaa 528

<210> 84
<211> 249
<212> DNA
<213> Homo sapien

<400> 84
acactgaagc agaaccggaa acaccagga actgttcaga aatctcagaa gaaatctgct 60
tctcttcgat ggaaagatat aattaacgat caaagagctc taagaaaatt gcaaagaagc 120
cttaatgttc aagctttaga aagatcagag caatttttct ctttcagtc aaactaagac 180
tctctgtatt taaatctctc tggggcaaga gggctagatt tcctcatttt gttatgagac 240
tagattggt 249

<210> 85
<211> 496
<212> DNA
<213> Homo sapien

<400> 85
actggccctc ggtgctggca aaggtgtagt tccactggcc gagggaatca agacatagtg 60
gtccttctgc taagccaagg gctgccacaa tgacacagta gccagatcct gcaattccaa 120
tgagagcagc caatacagaa gaaaacatcg cacatcgttt gccacagttt tcatggccac 180
agcagccaca gcagtcaccc tgttccagcc caatgaagac aaatgctggc aggagcatca 240
gcaggccacc tctacgatg ccagaaaaga accacacgaa gcggctgagg tggttttcgg 300
aggcatactt tgtttcccca ttgggaaagt aaagcaaaat attaaccgcg atgcacagga 360
gggcgagccc caccagagaa tgtccgatgc atcgtgcaca cttcccatag cacatggtgg 420
tctgctaggt tttctccccc ttctctttgt cttcagctca gtgatacccc aaattagatg 480
aaagtgtgcc cttctg 496

<210> 86
<211> 199
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(199)
<223> n = A,T,C or G

<400> 86
acagaaagag taagataaaa acatttaata tnattaaatc taatttgcaa aaattggtat 60
ctgacatttg ttgtgtgctc ttgcaaagag cgcataggac atttctgcag caatcaaaaa 120

ggtaaaatct ttttaaactc agatttcaag tttcctctaa tattccttct aatcctantc 180
cctggaaata ctttcaagc 199

<210> 87
<211> 436
<212> DNA
<213> Homo sapien

<400> 87
aacgttttga tttcatgaag gtgttctcaa atttaaagca cattttcagt aagaacaaaa 60
atatttaatg tttttatctt agacttaact tgatacattt gcatattact atggaagtta 120
ttcaccttgt ccctgttttt ctttaagata ttttaaaatc atagttatac tacagtcctt 180
ttttaaatgt atcctgatac attgtaaaat attttaattt cattgtggaa aataatgttg 240
gataaggaga tattttttcac tgttaacttt tagcccatgc attttcataa tttatttttt 300
tcacttgctg ctttatatga catatgtgac atttgattat ttaacacttg atgtgatctg 360
cataaaccga agttgcacaa ccctcctgct gaagataaaa ttgagggtta agataaagat 420
ttattttcat atttgt 436

<210> 88
<211> 596
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(596)
<223> n = A,T,C or G

<400> 88
acaaaagctg gtaatggacc aaagacttcc aaaatatatg tgtaatgacc tccagatttc 60
tttatagttg ttcccaattc agcataagac aaagctccaa atagtgcag gacccacac 120
accgtccaga tggtcagaga catgcccacg ctgcccgtgt tctggagcac gcccttagga 180
gagatgaaga ttctgtctcc aatgatgggtg ccaatgataa tggagactcc cctcagtaaa 240
gtgactttcc tcttcagctg cactttctcc tgcccagggtg gctccttggt gcccaggga 300
ggcagcctcc cgttaacatt tccctgcagg taacctcctt tggagatggt ggacacaaca 360
ggctttctga ccatagtagg gacacacggg ggaaaaataa aacagaggga aagaaaacaa 420
aactttcaac tttggtgtct cttggtgtta ctgatcgatg tcttcctctg ctttcagact 480
gtctctctca gcgctatagt gttcacagggt gaaaactcaa aggtgtgctt ttncttcac 540
agcgatctaa ttactactca gaaacacctg tgtatgcac gtgctctcaa ttcttc 596

<210> 89
<211> 435
<212> DNA
<213> Homo sapien

<400> 89
acacaagtca gtccaacagt tagtggttaat tactaataat atatgaaaac cctgccaca 60
caattgctgc tacatcacca atataattat taaccactgt cggaaaaaca cacataaatt 120
caggtaagac taaaagctgt ctcaaaaaa gaaaaaagaa atccaatgga tccactaatg 180
ctatcaaaag ggacatgcag gaatgtaaca tgacattttt agaaatgtgt gtttctaaaa 240
agaaaaaaa atacactaaa atgccagtgg actataattc attcaaaaca tctttagtgt 300
tccttcccaa agatcttgat ctgctcagta attgcttcac aagatctatc acagccatct 360
tttgagagct atggttaggc tggctcctct gtggtggtag gggcagtctt tttgaagctt 420
taagtatctg gtggt 435

<210> 90
<211> 344
<212> DNA

<213> Homo sapien

<400> 90

| | | | | | | |
|------------|------------|------------|------------|------------|-------------|-----|
| actcagcgcc | agcatcgccc | cacttgattt | tggagggatc | tcgctcctgg | aagatgggtga | 60 |
| tgggatttcc | attgatgaca | agcttcccgt | tctcagcctt | gacggtgcca | tggaatttgc | 120 |
| catgggtgga | atcatattgg | aacatgtaaa | ccatgtagtt | gaggtcaatg | aaggggtcat | 180 |
| tgatggcaac | aatatccact | ttaccagagt | taaaagcagc | cctggtgacc | aggcgcccaa | 240 |
| tacgaccaa | tccgttgact | ccgaccttca | ccttcccat | ggtgtctgag | cgatgtggct | 300 |
| cggctggcga | cgcaaaagaa | gatgcggctg | actgtcgaac | agga | | 344 |

<210> 91

<211> 371

<212> DNA

<213> Homo sapien

<400> 91

| | | | | | | |
|------------|------------|------------|------------|------------|-------------|-----|
| agcaatgcaa | aggacatctc | caatcatgac | atttaagaca | attctttatt | tctctgacag | 60 |
| tgacttcttg | aagtgcacat | ataataaata | aatagaaaat | atatctttgt | tcatgggtgat | 120 |
| gcctacaaga | aatgtttaca | tacaaacact | ctatacatct | aactcccgaa | aaaggaccag | 180 |
| ctatttcggc | aacagaaaaa | agacaagcat | ttcagaggag | cgttgctttc | cttaaagacc | 240 |
| taactcactt | aagtcttaca | aacagaaata | acaaggagga | caattttcta | agcaataaga | 300 |
| aaatttgtgc | taccaagaaa | atgcctagat | attggctctt | ggtgaatggt | ttaggaaaga | 360 |
| aacttttatg | t | | | | | 371 |

<210> 92

<211> 209

<212> DNA

<213> Homo sapien

<400> 92

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| acaacaaaag | atcaaaccce | tgtcccgatg | ttaacttttt | aacttaaaag | aatgccagaa | 60 |
| aaccagatc | aacactttcc | agctacgagc | cgccacaaa | ggccaccaa | aggccagtca | 120 |
| gactcgtgca | gatcttattt | tttaatagta | gtaaccacaa | tacacagctc | tttaagctg | 180 |
| ttcatattct | tccccatta | aacaccagt | | | | 209 |

<210> 93

<211> 176

<212> DNA

<213> Homo sapien

<400> 93

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| actccctggt | ttgagaaact | ttcttgaaga | acaccatagc | atgctgggtg | tagttgggtc | 60 |
| tcaccactcg | gacgaggtaa | ctcgttaatc | cagggtaact | cttaatggtg | cccagcgtga | 120 |
| actcgccggg | ctggcaacct | ggaacaaaag | tcctgatcca | gtagtcacac | ttcttt | 176 |

<210> 94

<211> 494

<212> DNA

<213> Homo sapien

<220>

<221> misc_feature

<222> (1)...(494)

<223> n = A,T,C or G

<400> 94

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| aatggaaat | ttaantgaca | tcctanaggt | agagaaaccg | nggagatcnc | ttttctcaga | 60 |
| ctcaccaact | tttaatggga | tttcatgggg | tttgggtgtg | ctgatagggt | aaggggaggc | 120 |

| | | | | | | |
|------------|------------|------------|------------|-------------|------------|-----|
| tgctttctgc | ccttctcccc | actcccatct | gatttactta | attcagtcctc | agctgctgaa | 180 |
| atttggaag | gaccaaattg | ctttacagtt | tttttctttg | cgtagtatct | tgaaatcctg | 240 |
| gaaaattcta | tggaatagtt | ctgtatatag | ggcacaagta | aaggcattgt | ccaaagttta | 300 |
| tttatttatt | tattacccta | agaatgcttt | gccataacca | catttaattg | gaaaaacggc | 360 |
| annatcacag | atgtaaatta | nctcaccana | tttactgngc | ctgaactcat | tctcttcttg | 420 |
| ctatatgatt | tagcaagttc | tagaaggnc | ccaagacaat | aattacattg | gcacaatgta | 480 |
| tacttcagng | ctca | | | | | 494 |

<210> 95
 <211> 260
 <212> DNA
 <213> Homo sapien

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| <400> 95 | | | | | | |
| cgcggcgagg | tacgggcttt | ccatctagtt | gccagcttag | atctgggggt | ggtaacccac | 60 |
| tgactttgca | gtccattctg | cagagttttc | cttcttgaac | agtcagatct | ccaggagcct | 120 |
| gcaagaagtg | aggtctgaag | aatcgctcct | gaattgggtc | atcttcgtct | ccactgtccc | 180 |
| ttgatctaga | acgaggcctt | ctgacatgag | gatggcctga | gggagaccgg | ggactccgac | 240 |
| ctctttgggt | gacagcctgt | | | | | 260 |

<210> 96
 <211> 438
 <212> DNA
 <213> Homo sapien

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| <400> 96 | | | | | | |
| accagttcct | gtttatatac | agtagtggtt | tgggcacacc | taaggtcgat | ctgtgttgta | 60 |
| tttaaaaatc | taatttcttt | atttgtgtgg | ccttctagac | aaacgaagg | gacccagagg | 120 |
| aaacccctg | acagatctct | ggatgatcct | ccttgaatcc | tgggcagttt | ggtctctcct | 180 |
| tgctgtgctc | ctgtggcact | aaactccttt | tgattgggtc | tttctttcct | tcccagctag | 240 |
| actaagccc | tcatgggcag | gtaatgaaga | ttgaaaactt | tttctgttct | tccagtgtga | 300 |
| gcacattcct | cctacatggt | agatgtgcaa | tagatgtttt | taaaattgga | gaatgaaaat | 360 |
| aaaagaagaa | aatcacaatt | tcttatcaag | ttgtagcttg | gtatcataca | caattgcatt | 420 |
| ctgaggaatt | aaggtggt | | | | | 438 |

<210> 97
 <211> 454
 <212> DNA
 <213> Homo sapien

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| <400> 97 | | | | | | |
| gagtaattcc | cctccagcac | tagagaccgc | tcagtgcctc | tactagatga | actcagtaac | 60 |
| gccttgagct | gggttgattg | aggatgtgtg | aaaagctcac | agagctcgat | gcctgctgct | 120 |
| atttcacggc | aatgagcctt | tttctttcta | cactgaagat | tttcttctta | tttaattgtg | 180 |
| tttatttttg | gctcagaaat | aattgctctg | ttgaaaataa | tcctttgtca | gaaaagaagg | 240 |
| tagctaccac | atcattttga | aaggaccatg | agcaactata | agcaaagcca | taagaagtgg | 300 |
| tttgatcgat | atattagggg | tagctcttga | ttttgttaac | attaagataa | ggtgactttt | 360 |
| tcccctgct | tttaggatta | aatcaaaga | tacttctata | tttttatcac | tatagatcat | 420 |
| agttattata | caatgtagtg | agtcctgcat | gggt | | | 454 |

<210> 98
 <211> 226
 <212> DNA
 <213> Homo sapien

<220>
 <221> misc_feature
 <222> (1)...(226)

<223> n = A, T, C or G

<400> 98

| | | | | | | |
|------------|------------|------------|------------|-------------|------------|-----|
| actaaatggt | ggtctaggag | cagctgggag | natagcaccg | ggcatatattt | ggaatggatg | 60 |
| aggtctggca | ccctgagcag | tccagcgagg | acttggtctt | agttgagcaa | tttggctagg | 120 |
| aggatagtat | gcagcacggt | tctgagtcgt | tgggatagct | gccatgaagt | aacctgaagg | 180 |
| aggtgctggc | tggtaggggt | tgattacagg | gttgggaaca | gctcgt | | 226 |

<210> 99

<211> 333

<212> DNA

<213> Homo sapien

<400> 99

| | | | | | | |
|-------------|-------------|-------------|------------|------------|------------|-----|
| actcatctag | acgttttaggt | atTTTTctgtg | gttgaggaag | ctcctctact | aaattcttaa | 60 |
| gaatatcttc | tggaatatac | tcatctggaa | aaagatgcaa | cctttccatc | attgttcttc | 120 |
| tgtgaagggt | ttttggcagc | atgccataaa | tagctagttt | tacaattgcc | actggatccc | 180 |
| tcaggatgaag | ctgagcagct | gttacttgtc | taaatccacc | tgggtagcca | gtatgcgaag | 240 |
| agtatacttt | ttgttcccat | ttgtttccag | aaaatgcaat | gtgtcttgtg | ttcattataa | 300 |
| caacatgatc | cccacagtca | ctcagtgcac | ggt | | | 333 |

<210> 100

<211> 417

<212> DNA

<213> Homo sapien

<400> 100

| | | | | | | |
|-------------|------------|------------|------------|------------|-------------|-----|
| accgccacat | cgctgacttg | gctggcaact | ctgaagtcac | cctgccagtc | ccggcggttca | 60 |
| atgtcatcaa | tggcggttct | catgctggca | acaagctggc | catgcaggag | ttcatgatcc | 120 |
| tcccagtcgg | tgcagcaaac | ttcagggaag | ccatgcgcac | tggagcagag | gtttaccaca | 180 |
| acctgaagaa | tgtcatcaag | gagaaatatg | ggaaagatgc | caccaatgtg | ggggatgaag | 240 |
| gcgggtttgc | tccaacatc | ctggagaata | aagaaggcct | ggagctgctg | aagactgcta | 300 |
| ttgggaaagc | tggctacact | gataagggtg | tcatcggtac | ggacgtagcg | gcctccgagt | 360 |
| tcttcagggtc | tgggaagtat | gacctggact | tcaagtctcc | cgatgacccc | agcaggt | 417 |

<210> 101

<211> 438

<212> DNA

<213> Homo sapien

<400> 101

| | | | | | | |
|------------|------------|-------------|------------|------------|------------|-----|
| acatatgttt | tttaagtaag | ttactttttac | cattagaata | aacctagaca | ctacagggac | 60 |
| aactctgggg | aacagggcgg | tctgccttaa | caacccttct | ctaggttgag | gaaggcaggt | 120 |
| atagttcact | gaaggatgtg | atgaggctgt | agtaagtctt | ctcatcatct | gttaatcctg | 180 |
| cgttgcctgg | tctcaccacc | acagctacgt | gcacatctgc | ttcctcagca | gcactggcct | 240 |
| ctcgagtaac | atctgtcaga | aacaaaatgt | tgttggttga | gcacccaatg | ctgtctgcaa | 300 |
| tctttcggtg | actttcactc | tctactttgt | gtccaatctt | ggtatcaaag | tgaccatcaa | 360 |
| caagctcaag | aatatctccc | tccgtagaat | gccgaataa | cagtttctgt | gcctccacac | 420 |
| tccctgagga | atagatgt | | | | | 438 |

<210> 102

<211> 466

<212> DNA

<213> Homo sapien

<400> 102

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| acttaaaaag | tggtttttct | atcttcaaag | tgctaaagaa | acaagtattc | aaaaagaaac | 60 |
| ttcaggctcg | tctacgaagt | tctgactgac | ttgaagtagt | gaaataccaa | gaatgcagtg | 120 |

gacaaattta aaaggccttc attagaataa agtatatctt aactacattt tgcaaagaaa 180
tgaagcaatg gttgcacaac cagtcagggc caagttagta acatacaact cagccatcag 240
cccacctctc cctcaaacta aactaatcta aatgtatttt tcagaaaatt tcctccatac 300
tccatgtatg tggtacatac atccaatcat atccatattt tggatcattt ttttctatat 360
tcatcagatt attgggttaaa atgcacagca agtagaaatg atccatttca aaattcttaa 420
tatctagcgt tctctgtaaa acaaaagctg acaacagttt tattgt 466

<210> 103

<211> 500

<212> DNA

<213> Homo sapien

<220>

<221> misc_feature

<222> (1)...(500)

<223> n = A,T,C or G

<400> 103

nggtgcagcg gagacagagg cggaagctgc agccctagag gtcctggctg aggtggcagg 60
catcttgga cctgtaggcc tgcaggagga ggcagaactg tcagccaaga tcctgggtga 120
gtttgtggtg gactctcaga agaaagacaa gctgctctgc agccagcttc aggtagcgga 180
tttctgcag aacatcctgg ctcaggagga cactgctaag ggtctcgacc ccttggcttc 240
tgaagacatg agccgacaga aggcaattgc agctaaggaa caatggaaag ggctgaaggc 300
cccctacagg gagcacgtag aggccatcaa aattggcctc accaaggccc tgactcagat 360
ggaggaagcc cagaggaaac ggacacaact ccgggaagcc tttgagcagc tccaggccaa 420
gaaacaaatg gccatggaga aacgcanagc agtccanaac cagtggcagc tacaacagga 480
gaagcatctg cagcatctgg 500

<210> 104

<211> 422

<212> DNA

<213> Homo sapien

<400> 104

tggttctagg agatatcaat accaaaccaa agaaagaaaa tattatagct tttgaggaaa 60
tcatgaagtc tgtatggctc aatgatttcc tgaagatgat aaagagcaag atattgcaga 120
taaaatgaaa gaagatgaac catggcgaat aacagataat gagcttgaac tttataagac 180
caagacatac cggcagatca ggttaaataa gttattaaag gaacattcaa gcacagctaa 240
tattattgtc atgagtctcc cagttgcacg aaaagggtgct gtgtctagtgt ctctctacat 300
ggcatggtta gaagctctat ctaaggacct accaccaatc ctectagtgc gtgggaatca 360
tcagagtgtc cttaccttct attcataaat gttctataca gtggacagcc ctccagaatg 420
gt 422

<210> 105

<211> 326

<212> DNA

<213> Homo sapien

<400> 105

acgaagtagg tccaaagttg ttgaccgtat ttacagtctc tacaaactta cagctcataa 60
acataaaatg aatactgaaa gaatacttta caagcaaaag aagaattctt ctataagcat 120
tcctttttatc ccagaaacac ctgtaaggac cagaatagtt tcaagactta agccagattg 180
ggtttttgaga agagataaca tggaagaaat cacaaatccc ctgcaagcta ttcaaattgt 240
gatggatacg cttggcattc cttattagta aatgtaaaca ttttcagtat gtatagtgtg 300
aagaaatatt aaagccaatc atgagt 326

<210> 106

<211> 543

<212> DNA

<213> Homo sapien

<400> 106

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| acttgtaatt | agcacttggt | gaaagctgga | aggaagataa | ataacactaa | actatgctat | 60 |
| ttgatttttc | ttcttgaaag | agtaagggtt | acctgttaca | ttttcaagtt | aattcatgta | 120 |
| aaaaatgata | gtgattttga | tgtaatttat | ctcttgtttg | aatctgtcat | tcaaaggcca | 180 |
| ataatttaag | ttgctatcag | ctgatattag | tagctttgca | accctgatag | agtaaataaa | 240 |
| ttttatgggt | gggtgccaaa | tactgctgtg | aatctatttg | tatagtatcc | atgaatgaat | 300 |
| ttatggaaat | agatatttgt | gcagctcaat | ttatgcagag | attaaatgac | atcataatac | 360 |
| tggaatgaaa | cttgcataga | attctgatta | aatagtgggt | ctgtttcaca | tgtgcagttt | 420 |
| gaagtattta | aataaccact | cctttcacag | tttattttct | tctcaagcgt | tttcaagatc | 480 |
| tagcatgtgg | attttaaaag | atttgccctc | attaacaaga | ataacattta | aaggagattg | 540 |
| ttt | | | | | | 543 |

<210> 107

<211> 244

<212> DNA

<213> Homo sapien

<400> 107

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| acaaaaatgg | ttataaaatg | gttgaagcaa | ctagaagcgt | gacaggtata | atacatataa | 60 |
| atacaaccaa | aattcaattc | aatgcaaagt | tgaatgacat | catattgcac | caaaatttat | 120 |
| tccatacaaa | agcacatgca | tcaagagttt | tcataagatg | aaaacaaaca | cacttacttc | 180 |
| atagcatctt | accacttact | tacacaaata | gcccataaac | accatctggc | attgtgattg | 240 |
| cagt | | | | | | 244 |

<210> 108

<211> 511

<212> DNA

<213> Homo sapien

<400> 108

| | | | | | | |
|-------------|------------|------------|------------|------------|------------|-----|
| acttcatgtg | atttgtcaac | catagtttat | cagagattat | ggacttaatt | gattggtata | 60 |
| ttagtgacat | caacttgaca | caagattaga | caaaaaattc | cttacaaaaa | tactgtgtaa | 120 |
| ctattttctca | aacttgtggg | atTTTTcaaa | agctcagtat | atgaatcatc | atactgtttg | 180 |
| aaattgctaa | tgacagagta | agtaacacta | atattggtea | ttgatcttcg | ttcatgaatt | 240 |
| agtctacaga | aaaaaaatgt | tctgtaaaat | tagtctgttg | aaaatgtttt | ccaacaatg | 300 |
| ttactttgaa | aattgagttt | atgtttgacc | taaatgggct | aaaattacat | tagataaact | 360 |
| aaaattctgt | ccgtgtaact | ataaattttg | tgaatgcatt | ttcctgggtg | ttgaaaaaga | 420 |
| agggggggag | aattccaggt | gccttaatat | aaagtttgaa | gcttcatcca | ccaaagttaa | 480 |
| atagagctat | ttaaaaatgc | actttatttg | t | | | 511 |

<210> 109

<211> 652

<212> DNA

<213> Homo sapien

<220>

<221> misc_feature

<222> (1) ... (652)

<223> n = A,T,C or G

<400> 109

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| acaccccaaa | ctctcccttg | ggagcctcaa | tggcagtata | tgtggctcct | ggaggaactt | 60 |
| ggtagccctc | agtatacaac | ttaaagtgat | gaatcagtga | ctccatggaa | gtcttcatct | 120 |
| ctgctcgctt | aggtggagac | actttggcat | catcaacctt | gatctcccca | ggaggcatct | 180 |
| tgtttagaca | ctgtgcgata | attctcaggg | actggcgcat | ctctccacc | cggcacaggt | 240 |

| | | | | | | |
|------------|------------|-------------|------------|------------|------------|-----|
| acctatcata | gcagtcccct | cgagaaccaa | caggaacatc | aaactcaacc | tggtcgtaaa | 300 |
| catcataggg | ctgggtcttc | cgcaggtccc | actggatgcc | tgagccccga | agcatcactc | 360 |
| cactaaaacc | atagttaagt | gcttcttctg | ctgttacaac | cccaatgtca | attgtccgat | 420 |
| ttcgccagat | cctattgttg | gtcagcaact | cctccaactc | atcaagccga | agagagaagt | 480 |
| tcttagaaaa | ctgataaatg | tcattccataa | gcccgaaggg | taggtcctgg | tgcactcctc | 540 |
| ctggccggat | ataagcagca | tgcatcggg | ctncagacac | ttcgctcgta | gaactcaaac | 600 |
| atcttctncc | tttcttcaaa | cagccagaag | aaaggggtca | tgggcccaag | gt | 652 |

<210> 110

<211> 96

<212> DNA

<213> Homo sapien

<400> 110

| | | | | | | |
|------------|------------|------------|------------|------------|------------|----|
| acacattgag | tattccacag | atatacatgg | tttaatatgt | ggtatccatg | gggtatgatt | 60 |
| ctaccacagc | cttgtaagtg | ctccaaacct | taaagt | | | 96 |

<210> 111

<211> 371

<212> DNA

<213> Homo sapien

<400> 111

| | | | | | | |
|------------|------------|------------|------------|-------------|------------|-----|
| acatagcagc | ttcataacag | tttacttttt | taatataaag | atTTTTcaat | ttacacttgt | 60 |
| aggagtagaa | aaaactaata | tgctaagtct | gtaagctacg | cagcaaaaat | aatgatctta | 120 |
| atgaagccag | aattctgtga | aaatgtgcac | cacactgcat | atatagtagc | tgagtaaagt | 180 |
| taaaccatgt | gcttattaac | tcttctatat | aaaatattga | accccccaagt | ctcacacatt | 240 |
| gcctcctatg | tccacatcac | ttttctgaag | acagcctcat | gctttaagcc | aatatatatt | 300 |
| tgctatttga | aaaagttctc | atcctcatta | ctaaaaatgt | ttctgtaaag | gccttagaca | 360 |
| tttttttcag | t | | | | | 371 |

<210> 112

<211> 406

<212> DNA

<213> Homo sapien

<400> 112

| | | | | | | |
|------------|------------|-------------|-------------|------------|-------------|-----|
| caggtacagt | aatacacggc | tgtgtcctcg | gttttcaggc | tgctcatttg | cagaaacaac | 60 |
| gtgtcttctg | aatcatctct | tgagatgggtg | aatctgcctt | gcacgggtgc | agcgtagtct | 120 |
| gttggtccac | catcagttgt | gcttttaata | cggccaaccc | actccagccc | cttccttgga | 180 |
| gcctggcgga | cccagctcat | ccaggcgtca | ctgaaagtga | atccagaggc | tgacacaggag | 240 |
| agtgttaggg | accccccagg | ctttactaag | cctccccccag | actccaccag | ctgcacctca | 300 |
| cactggacac | catttaaaat | agcagcaagg | aaaatccagc | tcagcccaaa | ctccatgggtg | 360 |
| agtctctctg | gttcagtcct | gatcactgaa | tgaaaacact | tgaggaa | | 406 |

<210> 113

<211> 492

<212> DNA

<213> Homo sapien

<400> 113

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| accatcccca | gaagtgtctg | gtgccaggca | ctgatccagc | agctcttcca | caatggatga | 60 |
| caataaccga | agctcccat | tttcatcacg | ctggctgatc | tttgattgaa | tgaaatctac | 120 |
| aacttcctgg | ctgctcatca | cattccagat | gccatcacag | gcaatgacca | tgaattcatg | 180 |
| gtcgtcagtg | agagtcagca | ccttgatgtc | aggaagggct | gaaatcatct | gttcctcagg | 240 |
| tggcaggttc | ttgtttctct | tgtagaagtg | gtccccaatg | gctctggaga | ggttgaggcc | 300 |
| cccgttgact | cgcccatcca | tggtgacctt | gccaccagca | ttcttgatgc | gtgctagttc | 360 |
| tacttcatcc | tctggtttgt | gatcatagga | catgtctaaa | gctttgccag | cctcagatac | 420 |

cacacagcga gagtctcctg cgttggctac aatcaactgc ttctctcgta tcagggccac 480
caccgctggt gt 492

<210> 114
<211> 234
<212> DNA
<213> Homo sapien

<400> 114
acctcagtgc aaaagttagt tgaactgggt cattcatctc tatggtaaca gcttcctcct 60
ctttatcgac attacttgtc tgtgacaatt taatgtttcc atttccaagt tctccacttg 120
cagaaaattt cactccgtct tttgcacagg aaattacaac agcatctcca atatggctga 180
gatctcggca tatacgtgca aattcaccag aaggcatctt tactacacag ctgt 234

<210> 115
<211> 368
<212> DNA
<213> Homo sapien

<400> 115
cctgggggtgg gatcagagga tctggcgtgg catcccgtag ccagtcatgc ctgcctgaga 60
cgccccgagg ttggtgcca tctgtaaccc gatcacgttc ttgccctctt gcagctggtt 120
atccgagaag ttccgaggat tctccttgga tttcttaggg aaccagttgg gatccccaga 180
gaagagccca tcatctcggg ctactgccag cccaccaga ttcatcagcg tccgctgcac 240
acaggccatg ttctttcctt cccagaggtc cacagtttgg aagatgtcag tgggtttaat 300
gccatagcgc tcagctgctt gcaggaactg agagatctgc tccatctgct tgaaggccat 360
ggtggagg 368

<210> 116
<211> 487
<212> DNA
<213> Homo sapien

<400> 116
ggatttttta ttgtgttttc cacatagata aaaaaataag gctttttgat gaaaagaatc 60
cattacaaag tcaaaaatcc attacaatta taattgaatc agtaacaaaa tttagcttta 120
aatgagtcaa gtattctgca tttgaaattt aatatcacia acattcaaga ttagtgaatt 180
ttggtaagaa aaaaatacta gaagaaagga aaaggacacc ttttcaacag atagtaattt 240
ataaaaattt ttttaaaagt gctttgggaa aacacacagt atcattactt aagaaaagtc 300
atttaaggaa gacttaagtg cttcaagtgg agtgatttac agactaaaaa atgttttaaa 360
atttgccaag aaatttaagt gttaaaaata ctcttctcct tattcagttt catgtttaag 420
gaaacatttg acagacaagt aaaccaaacg caaaaaaaag ttcacctgca ttttaaaacta 480
ataaatt 487

<210> 117
<211> 430
<212> DNA
<213> Homo sapien

<400> 117
gttttacttg ttgatttttg gatgcatgct gggggaggaa agcatattgt ttgtagtcac 60
cctagagtgc taaggatat tattccccag taattctctc aagggtggga tatgcaaaac 120
ataatctcta aattcttcaa tactaagaaa tacctttggt ttacccttaa aatcaaatgc 180
cattttggct ggatatagga ttctaggatt aaagcctttt tccagcagaa ctttgaagac 240
attgctccat ttacttctag catccagtgt gtccagtgat aagtctgctg tcaacctgat 300
tcttgctcct tggtaggtaa tttctcttct ctctctagaa gcccttatta ttttctcttt 360
atcactagaa ttccaaaatt tcaccaagat gtgtctagga gtcagtctct tttcatcaat 420
tttactaggt 430

<210> 118
<211> 305
<212> DNA
<213> Homo sapien

<400> 118
cctgctagaa tcactgccgc tgtgctttcg tggaaatgac agttccttgt tttttttgtt 60
tctgtttttg ttttacatta gtcattggac cacagccatt caggaactac cccctgcccc 120
acaaagaaat gaacagttgt agggagaccc agcagcacct ttcctccaca caccttcatt 180
ttgaagttcg ggtttttgtg ttaagttaat ctgtacattc tgtttgccat tgttacttgt 240
actatacatc tgtatatagt gtacggcaaa agagtattaa tccactatct ctagtgcttg 300
acttt 305

<210> 119
<211> 367
<212> DNA
<213> Homo sapien

<400> 119
cggtagaaga catcaaagtg aagtaaagcc caagtgttct ttagcttttt ataatactgt 60
ctaaatagtg accatctcat gggcattgtt ttcttctctg ctttgtctgt gttttgagtc 120
tgctttcttt tgtcttttaa acctgatttt taagtcttc tgaactgtag aaatagctat 180
ctgatcactt cagcgttaaag cagtgtgttt attaaccatc cattaagcta aaactagagc 240
agtttgattt aaaagtgtca ctcttcctcc ttttctactt tcagtagata tgagatagag 300
cataattatc tgttttatct tagttttata cataatttac catcagatag aactttatgg 360
ttctagt 367

<210> 120
<211> 401
<212> DNA
<213> Homo sapien

<400> 120
acaggtaaat aaaagatcac cttgaattaa actggatctc cttaagggca tagtatagtt 60
tcagtttcat tacctattac ataattagtt tcttacatac aaatattgac atatttggtc 120
tgtgcttcga agcctttgtg tctatgaagt ccacatcaat gcagctcata actggaagtc 180
actggggagt tctttgctgc tgctggggtt aacctgatca tgcattagag tctcctcagc 240
acctgttggt gctctgcaca cctctggggc atcgtcagtg tcaggatcca agccttcagg 300
gcagggaagt ttcagcaact cttecgcgag ctgagcagtg tgacgcttga gagctgctgc 360
atggtgagac atagtcctgc ctaccgctt atcactgctg t 401

<210> 121
<211> 176
<212> DNA
<213> Homo sapien

<400> 121
acagcccaga tgtgatattt ctacaggaag ttattccccc atattatagc tacctaaaga 60
agagatcaag taattatgag attattacag gtcatagaagg aggatatttc acagctataa 120
tgttgaagaa atcaagagtg aaattaaaaa gccaaagagat tattcctttt ccaagt 176

<210> 122
<211> 443
<212> DNA
<213> Homo sapien

<220>

<221> misc_feature
<222> (1)...(443)
<223> n = A,T,C or G

<400> 122

| | | | | | | |
|------------|------------|-------------|-------------|-------------|------------|-----|
| actgctgcca | gttccccacg | tggcccagcc | ccaccacacag | gctctcctgg | gcccaggaat | 60 |
| gtcctgcagg | agggaggagt | cggtttccaa | tgccagccgc | cctaacaacc | caggaactca | 120 |
| gctcaactgg | ttacagacct | cgagttttca | gcccatgtta | cttgaaggag | aagcagttct | 180 |
| tgggctttac | cacctgccac | ctggggccaga | gttctcttat | ccttataccta | agagtcttta | 240 |
| agactcaaag | aagaaaaggt | cttgtctgat | gtataatctt | aaaataaacc | cacacttagc | 300 |
| cacctcaaat | cctttctgaa | attatgtaag | atgaaaactt | aaatgcctta | tagataccaa | 360 |
| gtatctcctc | acaatattga | attccatgaa | accacttatc | tttgcattga | atgaagcatc | 420 |
| cacaaaacca | tttcaagctg | aan | | | | 443 |

<210> 123
<211> 520
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(520)
<223> n = A,T,C or G

<400> 123

| | | | | | | |
|------------|------------|------------|-------------|------------|------------|-----|
| actgtatatt | ngaagattgc | taagataatg | gatttttaagt | gatctcacca | caaaaaaaga | 60 |
| agtatataag | gtattagata | tgtaatttag | cttgatttag | ttattctaca | aggtatccat | 120 |
| atatcaaaac | atcatgttat | ataccatgaa | tatagacagt | ttctgtcagt | taaaagtaaa | 180 |
| taaaaatttt | aaaaaattat | caattcgtta | attttaccac | gttggggcaa | aagcctttta | 240 |
| acagtccang | aaatatttaa | agctagtcaa | cagcttctac | agagatgaag | aacattntgt | 300 |
| cctaaggggt | ttctgtaggg | atcaccccca | tctctagact | tctacctggg | aaacacgcct | 360 |
| tccactgggt | gatgaganta | aggtgatgga | ctgtcgatca | actaggncca | aggcctgggt | 420 |
| agctgatgag | ccaaagagaa | acttcagcct | gtgaaataaa | aacacttcag | attagaangc | 480 |
| ctgatttctc | aagtcacctc | agtaacttgc | ccaaggatcc | | | 520 |

<210> 124
<211> 406
<212> DNA
<213> Homo sapien

<400> 124

| | | | | | | |
|------------|------------|------------|------------|------------|-------------|-----|
| actaaaaatc | aattggatga | actaaatcca | aaacatgaca | ctgtaggcag | cagttttaag | 60 |
| tcttattttt | actgtttata | tatttgaatg | ctgctacaac | agatgatctt | catccctgaa | 120 |
| gttttcagct | aaacttggtt | tcctagaata | gactgttaac | tttcaaaatt | tttattgggtg | 180 |
| aaatggaaat | actgtttttc | cttgtgaatg | aattttcata | tttgtaagtg | ctaagtttat | 240 |
| aattcaggtt | tgatcaaggt | gtgaataact | gaagaaaata | acttgctggc | tatataggaa | 300 |
| aatgctgtgg | aatgaactg | tgtatatact | tctgggagga | acaaatttaa | tcatttcttc | 360 |
| tgtaagcac | taatcagtat | aagtgcact | cctggttctg | tacctg | | 406 |

<210> 125
<211> 413
<212> DNA
<213> Homo sapien

<400> 125

| | | | | | | |
|------------|------------|------------|-------------|-------------|------------|-----|
| gttttctttg | aatgatttct | ttttttcact | gtaagacact | ccttttaaata | atgcctatct | 60 |
| ttaacttttt | aagactatct | ggaaaaatgc | agtgctctcag | ctgtccccag | ggaaattaag | 120 |
| tggaattcaa | ctaagatctg | ttaataagat | gtcagaataa | ctaataattt | tattaggaaa | 180 |

| | | | | | | |
|------------|-------------|------------|-------------|------------|------------|-----|
| aaatcatggt | ttaaattttca | aatgacact | tattttgtcaa | gtaatatgat | cttggaaaat | 240 |
| tttaaagaaa | aataatccta | cttataaact | actttttttat | aattgttttc | agaaaaaaag | 300 |
| tttacagtct | taaggaaaat | attcaggtct | atcatatggt | ttgacagatt | ttttaaaagt | 360 |
| tatttttggg | aaggtcttct | tttagaaaaa | aattaatctc | aagggttttt | tgt | 413 |

<210> 126

<211> 655

<212> DNA

<213> Homo sapien

<400> 126

| | | | | | | |
|-------------|------------|------------|------------|------------|------------|-----|
| gtattctata | gtgtcaccta | aatagcttgg | cgtaatcatg | gtcatagctg | tttcctgtgt | 60 |
| gaaattgtta | tccgctcaca | attccacaca | acatacgagc | cggaagcata | aagtgtaaag | 120 |
| cctgggggtgc | ctaagtgtg | agctaactca | cattaattgc | gttgcgctca | ctgcccgtt | 180 |
| tccagtcggg | aaacctgtcg | tgccagctgc | attaatgaat | cggccaacgc | gcggggagag | 240 |
| gcggtttgcg | tattgggcgc | tcttcgctt | cctcgctcac | tgactcgctg | cgctcggctg | 300 |
| ttcggctgcg | gcgagcggta | tcagctcact | caaaggcgg | aatacggta | tccacagaat | 360 |
| caggggataa | cgcaggaaag | aacatgtgag | caaaaggcca | gcaaaaggcc | aggaaccgta | 420 |
| aaaaggccgc | gttgctggcg | tttttccata | ggctccgccc | ccctgacgag | catcacaaaa | 480 |
| atcgacgctc | aagtcagagg | tggcgaaacc | cgacaggact | ataaagatac | caggcgtttc | 540 |
| cccctggaag | ctccctcgtg | cgctctcctg | ttccgaccct | gccgcttacc | ggatacctgt | 600 |
| ccgcctttct | cccttcggga | agcgtggcgc | tttctcatag | cttcacgctt | gtaag | 655 |

<210> 127

<211> 442

<212> DNA

<213> Homo sapien

<220>

<221> misc_feature

<222> (1)...(442)

<223> n = A,T,C or G

<400> 127

| | | | | | | |
|-------------|------------|-------------|------------|------------|------------|-----|
| accttatggt | ccttgaaagg | aagactcaat | acttccagga | gtcaaagtta | atttgaatga | 60 |
| aatggaaga | gaacaagttg | acaataattt | gaagcaattc | atgcttctag | ggctgaatga | 120 |
| cgttttagatc | agacacagag | tgactgagcc | aatcaacagg | catgtagtgt | gatctttccc | 180 |
| accacagtga | acagagggat | tctttgtcca | aggcaggctt | gcagctcggg | ccagcttgag | 240 |
| catttgatca | ggatttgatg | cttcaaagat | gaccactct | ctgtaaactc | attaccaaa | 300 |
| caaaatgcaa | tgatctcttc | catttggtgga | acataccacc | aacacaaacc | acgcgtggct | 360 |
| ttgcctcctg | ttcactccat | tttcaaggct | agagaaagtt | caagtccaaa | acaacagtta | 420 |
| aggntaaaac | gctaaacctc | aa | | | | 442 |

<210> 128

<211> 447

<212> DNA

<213> Homo sapien

<400> 128

| | | | | | | |
|------------|-------------|-------------|------------|------------|------------|-----|
| gtaaaatctg | atggtgggtta | aatgacgatg | tttaggtttt | gataaattta | gattttatac | 60 |
| acatgataga | gcatgtatct | gtatttttta | aaataaagac | agagaactta | tgtttagaac | 120 |
| aagagaagcc | atttggtaga | aataaagaag | gagattgggg | aaggagatga | gaatgagtca | 180 |
| gagagatagc | atttaaaact | tgaaatcagg | cacaacaatt | agtatgtcat | gatataaaca | 240 |
| gtattgagat | aaaattttac | cacttctctt | ccctttaata | aattgtcaaa | ggataaagtt | 300 |
| tcctgtttga | aaatatattt | tactgggtatt | gtgcttttct | catatcacag | attggtaaag | 360 |
| aatcatttta | agtccaagac | tcttatttta | catattctgc | aattaaaggt | cctatgaggc | 420 |
| tacctgccga | ctgctgacat | gtagtgt | | | | 447 |

<210> 129
<211> 175
<212> DNA
<213> Homo sapien

<400> 129
ttcagacttt gttttagtagtc agccttggtt tggcttcaga ctttgtttgt cgtatttgag 60
gatataaata ttcatagaata gtttcccaag tctggagcga ccacataggg agaaaatgta 120
aatgtctcaa tttttgttca caaaagtata ttttatcaaa ttgctgtaag ctgtg 175

<210> 130
<211> 406
<212> DNA
<213> Homo sapien

<400> 130
acatttacat tcaagttgat aacactggtg gtttcatttc aatacaaatt atgctagaga 60
actgacattt cagacatggt catatatatg ctatttgaat tcctttatct tgatacagat 120
cttgattgtg aatctcttga tgatagatgt gcagctaatt tgtcccgaaa ctcatgaaga 180
taattgtatt gcttgatggt ctgtattgcc ccggatcctc ttaggtctcg caggctgtct 240
atggcttgct ctggtgatat tgtgtcagac aggtatagta ggagacaagc agctacaaga 300
caagatctcc caagtcctcc atagcagtgt attaagggtt ttcggttaatt ttttaaggcag 360
gttgtaagct cttccattat ttcacagcag ctggctatgt caggag 406

<210> 131
<211> 403
<212> DNA
<213> Homo sapien

<400> 131
accgcattac attatgcctg tgaaatgaaa aaccagtctc ttatccctct gctcttggaa 60
gcccgtagcag accccacaat aaagaataag catggtgaga gctcactgga tattgcacgg 120
agattaaaat tttcccagat tgaattaatg ctaaggaaaag cattgtaatc ctgtgacca 180
caccgatgga gatacagaaa aagttaacga ctggattcta tcttcatttt agacttttgg 240
tctgtgggcc atttaacctg gatgccacca ttttatgggg ataagatgc ttaccatggt 300
taatgttttg gaagagcttt ttatttatag cattgtttac tcagtcaagt tcaccatggc 360
cgtaatcctt ctaagggaaa cactaaagtt gttgtagtct cca 403

<210> 132
<211> 479
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(479)
<223> n = A,T,C or G

<400> 132
cgaggtacag ggggaccccc ttctcaacgg caccagcttt gcagacggca agggacaccc 60
ccagaatggc gttcgcacca aacttagatt tattttctgt tccatccatc tcgatcatca 120
gtttgtcaat cttctcttgt tctgtgacgt tcagtttctt gctaaccagg gcaggcgcaa 180
tagtttttatt gatgtgctca acagcctttg agacaccctt ccccatatag cgagtcttat 240
cattgtcccc gagctctagg gcctcataga taccagttga agcaccactg ggcacagcag 300
ctctgaagan accttttgag gtgaagagat caacctcaac agtgggattc ccgcgagagt 360
caaagatctc cctggcatgg atcttgagaa tagacatggt gaacttctag ccactgggtc 420
tcgtcgccta ggagaggaag cggaggggtgc tgcanacacc gaggtgaacg taaagcccg 479

<210> 133
 <211> 301
 <212> DNA
 <213> Homo sapien

<400> 133
 gtcttacagt gtgactcaga ctccctatct ggggatcggt taggttgctt caatctaact 60
 atcaaaggac acgccaagtg tgtggaattt gtcaagagct ttaacctgcc tatgctgatg 120
 ctgggaggcg gtgggttacac cattcgtaac gttgcccggt gctggacata tgagacagct 180
 gtggccctgg atacggagat ccctaataag cttccataca atgactactt tgaatacttt 240
 ggaccagatt tcaagctcca catcagtcct tccaatatga ctaaccagaa cacgaatgag 300
 t 301

<210> 134
 <211> 494
 <212> DNA
 <213> Homo sapien

<400> 134
 actaagtgtg tacgtatttt tgccactttt tcttcagatg attaaagtaa gtcaacagct 60
 tatttttagga aactgtaaaa gtaataggga aagagatttc actatttgct tcatcagtgg 120
 taggggggag gtgactgcaa ctgtgttagc agaaattcac agagaatggg gatttaaggt 180
 tagcagagaa acttggaag ttctgtgtta ggatcttgct ggcagaatta actttttgca 240
 aaagttttat acacagatat ttgtattaaa ttggagacca tagtcagaag actcagatca 300
 taattggctt atttttctat ttccgtaact attgtaattt ccacttttgt aataattttg 360
 atttaaaata taaatttatt tattttatct tttaatagtc aaaaatcttt gctgttgtag 420
 tctgcaacct ctaaaatgat tgtgttgctt ttaggattga tcagaagaaa cactccaaaa 480
 attgagatga aatg 494

<210> 135
 <211> 448
 <212> DNA
 <213> Homo sapien

<400> 135
 actgaactcc catcacaaca tcatcttctt ctaataactg taacacaaca ccttcaataa 60
 actttgcatt gggctctgcc atagctgctt tccggagact catgatgaat cttccgtgat 120
 ggaaagctct tccactctgc acttgattgt tttctgacag agggtaagga atctgaacct 180
 ctgatttgct ttcctgatca tgaatcatgt aaccatttac aacctgggca tcaagacctt 240
 ccactgtatc tccaagacca aggtctttga gaacatgata accacccggc tgcaggaatt 300
 ctccaactat tctgtcaggc tcttttaagt ctctctcaat gactgtcacc tttcttccat 360
 ctctggaaag cacagctgcc aaagcagagc caagcacgcc agctcccacg atgataactt 420
 ctgggtcatt ctgagaagat gttgatgt 448

<210> 136
 <211> 527
 <212> DNA
 <213> Homo sapien

<400> 136
 accatgggtg cagcaatttc ttccataact tcttggtaat ggtaattaaa agccatttca 60
 atgtccaaac caacaaactc agtttagatgt ctatgggtat tagagtcttc cgctctgaat 120
 actgggtccaa tagagaaaac cttctcaaaa tcagcacaaa tgcacatttg cttatatagc 180
 tgtggggact gagccaggta tgcattattt ttaaaatatg acacagtaaa aacattggct 240
 cctccttcac tggcagctga aataatttta ggagtttgga ttccacaaa acctttgtta 300
 attaaagttt ctcggaagag atggcagatg ccagactgga gacggaagac tgcctgacta 360
 gttgatgtcc taagatcaat gactctgttg tctaactctt tatcctgggt aacagtagct 420
 cttccttctt cttctccttc tgcctcaggc cgaacagcat catccagctg caggggcaga 480

cggggttcag ccaaactgat cacataaatc ttctgaacat gtaactc

527

<210> 137

<211> 275

<212> DNA

<213> Homo sapien

<220>

<221> misc_feature

<222> (1)...(275)

<223> n = A,T,C or G

<400> 137

| | | | | | | |
|------------|------------|-------------|------------|------------|------------|-----|
| acgacgagtc | gggcccctcc | atcgctccacc | gcanntgctt | ctaaacggac | tcagcagatg | 60 |
| cgtagcattt | gttgcattgg | ttaattgaga | atagaaattt | gcccctggca | aatgcacaca | 120 |
| cctcatgcta | gcctcacgaa | actggaataa | gccttcgaaa | agaaattgtc | cttgaagctt | 180 |
| gtatctgata | tcagcactgg | attgtagaac | ttgttgctga | ttttgacctt | gtattgaagt | 240 |
| taactgttcc | ccttggtatt | tgtttaatac | cctgt | | | 275 |

<210> 138

<211> 354

<212> DNA

<213> Homo sapien

<400> 138

| | | | | | | |
|------------|-------------|------------|-------------|-------------|------------|-----|
| caagctcaag | gtgttttctgt | caggaatgcc | agagctgcgg | ctgggocctea | atgaccgcgt | 60 |
| gctcttcgag | ctcactggcc | gcagcaagaa | caaatacagta | gagctggggg | atgtaaaatt | 120 |
| ccaccagtgc | gtgcggctct | ctcgctttga | caacgaccgc | accatctcct | tcaccccgcc | 180 |
| tgatggtgac | tttgagctca | tgatataccg | cctcagcacc | caggtcaagc | cactgatctg | 240 |
| gattgagtct | gtcattgaga | agttctccca | cagccgcgtg | gagatcatgg | tcaaggccaa | 300 |
| ggggcagttt | aagaaacagt | cagtggccaa | cgggtgtggag | atatctgtgc | ctgt | 354 |

<210> 139

<211> 527

<212> DNA

<213> Homo sapien

<400> 139

| | | | | | | |
|------------|-------------|------------|------------|------------|-------------|-----|
| acgaggaatg | acctctaggg | cctgggcaac | agccctgtat | ggccattggt | ccacaccagt | 60 |
| catggccttg | gatttttctg | tcaaggcatg | ggccacagcc | atctcggagg | ccccaccccc | 120 |
| tgacaccagc | tgaggggtcca | ggagaacatt | gcgacacact | tgcatggcat | cctggagggtt | 180 |
| gcgttctact | tccgagagaa | tctctttgct | agcccccccg | aggagaatgg | tgacggcctt | 240 |
| ggggtctttg | cagtcagtga | tgaaagtaaa | gtattcatct | ccaattttct | tgatttccaa | 300 |
| caggcctgct | cctgttccaa | catcatcttc | tctcagttcc | tctggctcgc | tgactatccg | 360 |
| ggccccacag | gctctagcaa | tgcgattatt | gtctgtcttc | cggactctgc | ggatggctgt | 420 |
| gatattggcc | cgcataaggt | agtgtctgag | taaatctgag | atgccctttt | cagtgatgac | 480 |
| cacatcgggc | ttcagttgga | taatgtcctc | acagagctgc | tggatgt | | 527 |

<210> 140

<211> 396

<212> DNA

<213> Homo sapien

<400> 140

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| acgccactgt | ctcttagata | taattatccc | caccctctgc | tcatttggtt | cccagattca | 60 |
| atacattgtc | aaagcctctt | ggtccttttt | taacatctca | cacttggtgc | attctctcca | 120 |
| ttcccataaa | cctcaacaac | tgctcaaagt | cctgcttgac | cccttggtgc | cagtctttga | 180 |
| aatctttctt | gcatatgact | gcctcattac | cttcctaaaa | tctagttcac | tcgcctactc | 240 |

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| aagaagacac | aggggcctac | tgtggtgtat | tagataagtt | cacatttctt | ctctttacta | 300 |
| atctttttta | cttcctttac | caccactccc | ttatataatt | ccatcatcct | aatagatctg | 360 |
| tttccctaca | catccctgcc | tctccacccc | acatgt | | | 396 |

<210> 141
 <211> 490
 <212> DNA
 <213> Homo sapien

<220>
 <221> misc_feature
 <222> (1)...(490)
 <223> n = A,T,C or G

| | | | | | | |
|-------------|-------------|------------|------------|------------|------------|-----|
| <400> 141 | | | | | | |
| acaaccagct | gtgctataag | aaagagggag | ggcctgacca | taactacacc | aaggagaaga | 60 |
| tcaagatcgt | agaggggaatc | tgcctcctgt | ctggggatga | tactgagtgg | gatgacctca | 120 |
| agcaactgcg | aagctcacgg | gggggcctcc | tccgggatca | tgtatgcatg | aagacagaca | 180 |
| cgggtgtccat | ccaggccagc | tctggctccc | tggatgacac | agagacggag | cagctgttac | 240 |
| gggaagagca | gtctgagtgt | agcagcgtcc | atactgcagc | cactccagaa | agacgaggct | 300 |
| ctctgccaga | cacgggctgg | aaacatgaac | gcaagctctc | ctcanagagc | caggtctaaa | 360 |
| tgcccacatt | ctcttntctgc | ctgctgttcc | ttctccttta | tggacgtcta | gtccttgtgc | 420 |
| tcgcttacac | cgcaggcccc | gcttctgtgt | gcttgtcctc | ctcctcctcc | caccccataa | 480 |
| ctgttcctaa | | | | | | 490 |

<210> 142
 <211> 511
 <212> DNA
 <213> Homo sapien

| | | | | | | |
|-------------|-------------|------------|------------|-------------|-------------|-----|
| <400> 142 | | | | | | |
| acatccagtc | tgtattttctt | acacaaaatt | acatctaatt | atttgacatg | aggtcatttg | 60 |
| ctatcataag | ccatcactag | gaacttctag | tctgtctcac | tcgattgagg | ctacaatgtt | 120 |
| gttaggtgct | atgaccacaa | tgaatacaac | agacagcctc | tcagctgtgc | tgcaaagtat | 180 |
| tcataaccaa | aagaccatat | ttcaaattaa | atcatagtag | cgaatgacat | accattttaca | 240 |
| tattacaatc | tgagcctctg | aaacaggggg | aacatataat | ggatatccaga | acatctttac | 300 |
| atcaaaaataa | cctatcacac | tacaaagttt | tcacttccaa | aaagtgtaac | agagttttaag | 360 |
| gcactggtaa | ctttgtccac | tgtagagat | taaaacttcc | aaagcaaatg | aaagaaccaa | 420 |
| tgttcacctt | taacgtgggg | aaagttggca | aaaagaaccc | caggaggaca | cccaaacctt | 480 |
| ctctgtgtcc | tctgtggaac | ctggcttttt | t | | | 511 |

<210> 143
 <211> 463
 <212> DNA
 <213> Homo sapien

<220>
 <221> misc_feature
 <222> (1)...(463)
 <223> n = A,T,C or G

| | | | | | | |
|------------|------------|-------------|------------|------------|------------|-----|
| <400> 143 | | | | | | |
| actgcagtga | ctcatcagag | tagaaggagt | attcaataag | tgggacttct | gtgtcggtta | 60 |
| attgggcata | tgctaaaaaa | gtgccgtttg | gagaccacca | cagagcagag | taggcactga | 120 |
| agacttcctc | ttcataaacc | cagtcagtta | ttccattata | tattatatct | tctttcccg | 180 |
| tccatgtgat | tctgtaactt | ggtaaatttg | gttcaatttt | aacataaatg | tcattgttcc | 240 |
| aaacatatgc | caatttatga | cccactgggtg | accatgtgac | ccactgtgtg | ttgtttggaa | 300 |
| tcctctcttc | tgtaatcagc | tgccttttat | ttaaatcata | aatgtcatat | gaagctgtgt | 360 |

aggaatgcct ccattgcttc acgtagttgt attctaagag aataaactgc ccatacangag 420
atattgaata atcattgata gaatgnccaa actcatcaaa tgt 463

<210> 144
<211> 297
<212> DNA
<213> Homo sapien

<400> 144
actcattaat attatatttgt ttgagaaag ccagaaatga ttctaagaaa taaacaataa 60
taataaaaga tgtaattaat atactgtatc ccttttaagc caaagcacac tttttacctc 120
aagactgttc tgacttttac attcttaatt tcctttgtcc aaaataggac cccattttta 180
atagagttca ttggaattga gttcataatc taaagtcact tttccccaca agatgttttc 240
atttcagtat ataaactgct aagcggcaaa tgactaagtc agttataaag aatttgt 297

<210> 145
<211> 356
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(356)
<223> n = A,T,C or G

<400> 145
actnctgcac ctcccttcagn aggaggncaa aggggaatgg cgacagctgc tcaatccttg 60
tgatggncac ctgccccacc atgtcgctg ctttgcgctc ccgggttgag gtcataatac 120
actttgccgg tgcagaanag aagccttttg acattttctg ggntctgagc tgcaaggcca 180
tcttctggga tcacccgctg gaannnggtn cctggaagca tctcatcaaa gctggatctg 240
gcctcggggn ggcncacaaan ggatttgggg gtgaagataa ttaacngctt ccggaatggc 300
agcnggatct ggcgtcgtaa cacgtggaag aagctgccac gagnggagca nttgac 356

<210> 146
<211> 355
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(355)
<223> n = A,T,C or G

<400> 146
acagttttgt tttctcgtaa ggggagcatc atagggttac tttataccag ttgtaacatt 60
ttcattgttt ttggttggtc ttttttcttt ttttaatggc agctaaagat atacagatta 120
ctgttaaatt gcagtccttt tttttttaaa natattttct tgagttattt aaaacatggt 180
aagcctggta ttttttaatc aaacaaaata tttatgaaan gggttttctc ttaattctgg 240
attcatcatg gctttctaata accaattgta atatttacia tattcaccaa aacttagaat 300
tttgcaaatg ctggaattct gccagtgttt ctttgctaag ccttgcatgc aaaat 355

<210> 147
<211> 209
<212> DNA
<213> Homo sapien

<400> 147
attttttact ttatatatga aaatgtcatg aaatttataa gcaataatgt attgatactc 60

| | | | | | | |
|------------|------------|-------------|-------------|------------|------------|-----|
| aaatttttaa | aaatttttaa | atttttaaaat | attttaatcaa | cttctattat | ttttcctctt | 120 |
| ctgggatgaa | ttaagtggca | aacttggcca | ttctaataatt | tactcactga | tagccaaatt | 180 |
| ttatagcgtc | tctatctaaa | gaagacagt | | | | 209 |

<210> 148
 <211> 445
 <212> DNA
 <213> Homo sapien

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| <400> 148 | | | | | | |
| actcccagca | aatcctctga | atactccaca | gactatgtta | cccagtccca | aggctattaa | 60 |
| ctcctgattg | ccatcaagtg | gataatcgta | tttgagggaa | tagacgctgg | caactgaaaa | 120 |
| ggccactgca | aatgcaacca | ttgcgatgcc | gaagcaatct | cctacggtgt | tttggaagt | 180 |
| ctccacgtca | ggtgtaatag | ggggctgaaa | tccaggattc | atgtcccaa | ccacagccac | 240 |
| tttaaacctg | tttttaaagt | cacagccgta | ggatacacct | gctgcaatca | cggtcataat | 300 |
| gaattcgatt | ggaatgggca | ctggaagttt | gtctttgaag | cgctgattta | tttctttaac | 360 |
| aatggataca | acaaaagga | caatcagagc | tgtcaccagg | tctgcaatat | tagtcttctc | 420 |
| tatttgtgag | aatacagagt | atagt | | | | 445 |

<210> 149
 <211> 585
 <212> DNA
 <213> Homo sapien

<220>
 <221> misc_feature
 <222> (1)...(585)
 <223> n = A,T,C or G

| | | | | | | |
|------------|------------|------------|------------|-------------|------------|-----|
| <400> 149 | | | | | | |
| actattaatg | agaacgaaat | acacattagg | aaaatggagc | catttcaatc | tagtggtttg | 60 |
| ggcaagatgg | ggaagagaag | gggaaacatt | ctagtttctg | gattacatta | ttatgcccct | 120 |
| cctgaaaagg | tggttgtcat | ttgcatttat | ttaaagcagg | taatatgcag | gaatgtaact | 180 |
| gaggattatc | ttcaggcaat | cagcaagata | tcctcctcat | ggtcccttta | gctctcaaaa | 240 |
| gcaatgaaat | cctcctgttc | tcatttttac | tgctgtgggt | gtgctgctga | acaatactat | 300 |
| cttctcaaat | tccatgccac | aaattcagca | ataacttttt | ggattgaatt | tagcaactac | 360 |
| tgtaattgga | tgctgatgtg | gacaaaatat | attgatttcg | atttcactcc | cgaatgtgat | 420 |
| tgccaccagc | tctttatatt | gctgctgtgg | tattttaaac | cagaagcttc | tttaaattat | 480 |
| gttgcaaact | gatctttgnt | tttatgtttt | ggtttggttt | tattttctaag | tgataagttt | 540 |
| gaaacacaca | gctttaaatg | atttttttat | tgtgggattt | tgggt | | 585 |

<210> 150
 <211> 508
 <212> DNA
 <213> Homo sapien

<220>
 <221> misc_feature
 <222> (1)...(508)
 <223> n = A,T,C or G

| | | | | | | |
|------------|-------------|------------|------------|-------------|------------|-----|
| <400> 150 | | | | | | |
| acaatgtctt | agaaagtctt | taagtcacat | accatgaatt | tttgcttcat | tactgaccat | 60 |
| atatgacctt | ggaggaactc | tttttttttt | ccttctactc | atttctgttt | ccacctaccc | 120 |
| tgactcaccg | tatttccagt | cttctacccc | tgcaattatc | ctagtccagc | aaagtcatth | 180 |
| ntttcaaaan | anacatcatg | tctgaaaata | attactggta | gtctaataatg | agccanagta | 240 |
| aacagctcct | catgggtcaat | gaacatgttc | aggaagcgat | caccttgatg | cttgaacca | 300 |
| accccanaca | gnggacaatt | ntactttgaa | atatccngna | atatttactg | ggggatccaa | 360 |

tttaaacttc tttnttctnt agcctttaaa ttacacaact ttgaactgac acggatctnt 420
tacaanaaac aatgcggcac tgaaggaana gatgattcct ttactcaaac ctgcaggaat 480
cagcctatta acaggcaggg gaaacggt 508

<210> 151
<211> 434
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(434)
<223> n = A,T,C or G

<400> 151
accatgaata aaagtgcatt tcaataccag ttttaacaac agcatatagg gcagacataa 60
aagaagacca cttccgaaac tagtgcaaga gattgagcat taggcacaaa gggagaaaaa 120
tgaaaagaat gaactttttg aaggaataag cattaagact agatgaccac attattatag 180
agacaaagct agcagcaaaa ttttaatcct tgatgatgta gctttcaaaa tttgcattct 240
ctcctatagt ctaccctata cgaacagctc ttcctatttt cctctttccg actgtgaagt 300
tactaaaatc ctaacactaa ttccatatat tctgtgtgcc aggcatttcc catgcttgct 360
atctaactcc cgggtaagca aatcttgnag taagaggcag tacctgectg gcggccggtc 420
aagggcgaat tctg 434

<210> 152
<211> 320
<212> DNA
<213> Homo sapien

<400> 152
actttgcaat catctttcct tttttcacat tggtaaaaat aagtggcatc cataggatca 60
tgatttttaa tttgttgccct ctgaagattt cactccatca agatctgccca atcttcaata 120
ttctggctaa atcttgggtat gtgggttttta aacagtcact ccgtttcaaa gtgtgtcttt 180
ccttatagaa tgtggaaatt atttctccat accttgtgat tttgacctga gtgctaagag 240
aatcactctc cttacctagt tatctacaaa tgttcattcc agaatgttt agttactgaa 300
ttgaatgaag acatctcagt 320

<210> 153
<211> 459
<212> DNA
<213> Homo sapien

<400> 153
acctcatttt tattagccat tatcttcatg ctggattcta atattctttt taatgggtgat 60
ctgttcaatg acagaaactt atagagagaa aattccttct caatttataa acaaaaattt 120
taaaagcagc atttttgatg tggtaggaag atatttatga caaaagcagc tactgcccta 180
aactggcaaa aacaacaaaa gaacaaattg ttatttaacc tttaaataac gagtctctat 240
ttgctataaa tctacaaata ttttaaatat atttctcct actgcaataa aaattaagat 300
aactctctgt ttaacagctt ttgaagagtt aattttataa ggaaataaaa aagattgact 360
tgcctcctga atgtccagtg ataaactgaa ccctaatttc cctacctcaa caacataaaa 420
atgatgtaaa gtggatcaaa gtatgtaaca agttaatat 459

<210> 154
<211> 503
<212> DNA
<213> Homo sapien

<400> 154

| | | | | | | |
|------------|-------------|-------------|-------------|------------|------------|-----|
| acacagcctt | gttgccatgt | ctgttggtggg | ccacaatcgc | cttgtccttc | tgaattatga | 60 |
| tttctggaaa | ctcctgggcc | aggtgagtc | cttgaatggt | gcacttaatg | tggagctgag | 120 |
| ctccttccat | gatcattccg | gtggggctga | tgtggaactt | gggtgtagag | aaggattccg | 180 |
| tcacggtgac | cagttcactc | ttggtagatt | ctgaggtctg | catatggatc | ccagaaatga | 240 |
| tcctagcttg | acgtcggaag | gataaaacgc | ggtcctgttc | ctcaacgggg | aattccagta | 300 |
| tcacaaaatt | ctgggtctcga | gaattcttct | ctcttttcag | cttgaccatt | ttttcattta | 360 |
| gttcaagttt | ttcaattgtg | aagtgtattg | gggccttttc | ctctgggaca | gaacagttga | 420 |
| ccctcacgat | cccaccttgg | atggcctctt | tcttgctccag | tgtcacctcg | ggactgggca | 480 |
| ctcctttcac | caacacctgg | tac | | | | 503 |

<210> 155

<211> 364

<212> DNA

<213> Homo sapien

<400> 155

| | | | | | | |
|------------|-------------|------------|------------|------------|------------|-----|
| actaaatata | gaacacttaa | caaagtccaa | tcttttgctg | agtgaaaatt | taacaattta | 60 |
| ctgagagaaa | agtaaatata | agaattttaa | gttcctttca | tacttgatca | tactataagc | 120 |
| attgccatca | tttcaatgca | catatatatt | taaaaaacia | ttttctctct | caaactcata | 180 |
| ttaaataact | ggatttttaa | acattttccc | catccacaca | aaaaagatat | gtgggttcta | 240 |
| attattcttt | gctattttaat | aatgctacct | ttgaagattt | ctacataata | taaacattcc | 300 |
| aattctgaag | caaagtattt | cagcattttt | caaaagtctc | taatatatct | tttgtttgta | 360 |
| gcgt | | | | | | 364 |

<210> 156

<211> 452

<212> DNA

<213> Homo sapien

<400> 156

| | | | | | | |
|------------|------------|------------|-------------|------------|------------|-----|
| acatatatgt | atattatacc | aatagctagt | aattttcaaaa | aaaacattga | cttgagtgtt | 60 |
| agataaccat | tctctaaatt | cagtttttga | tgtttcaaga | aacccaaaag | cctgtctttt | 120 |
| cacctacaga | ccctttgtgc | acgtggcaaa | tcacctctga | aaggcaaaaa | actaactgga | 180 |
| ttctcttcat | ttgttcaaaa | aagagaagaa | agctttaaag | atatgcctat | aaataaaaga | 240 |
| aaattaggtt | gctatattat | gattgtgcaa | taagtattaa | tttcattgaa | gtttgacctt | 300 |
| gttccatgta | ttagatgact | aagacattta | actcttaggg | atgttgaaag | cgcaccacaa | 360 |
| aacataagta | atcaataaag | taatgtttga | agacttttag | tatatactgc | ttattcaggt | 420 |
| aattaattat | tttgtaaata | ctaatagcac | at | | | 452 |

<210> 157

<211> 224

<212> DNA

<213> Homo sapien

<400> 157

| | | | | | | |
|------------|------------|------------|-------------|------------|------------|-----|
| acatgaacag | caggctgttg | cattgtaact | tgtggctgtg | cattaagatg | ttgctgagga | 60 |
| ttgcgaactc | ctgcagcata | tttatactgt | ggaacgggtg | ggacagcagg | agtagctgca | 120 |
| gcggctgcag | ctgcaggacg | tggacccatt | gtctgtgttg | atgtgttagc | aacacgctgt | 180 |
| gttgacatga | ctcgtggaac | ctgtgaagaa | gctggctctca | tagt | | 224 |

<210> 158

<211> 623

<212> DNA

<213> Homo sapien

<400> 158

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| acacatttca | ttatgctgcc | ttttctctta | tgattaaaac | tttagccctc | attcgaggtt | 60 |
| tccaatgggt | acttttagtg | gaggagttec | ctagctttta | aaaaaccact | tttctcttaa | 120 |

| | | | | | | |
|------------|------------|-------------|------------|------------|------------|-----|
| gattccatta | tttattgaaa | gaagtctttc | tagaaatggt | aaggaggatt | ttaaatgaac | 180 |
| acattcaatt | aaaaaaaaaa | tcacgtattg | aacatctacc | aagcatctgg | actcttcgga | 240 |
| acctagtaaa | atgaaaaaat | ccagttttta | caacagtaac | ttcattctgc | gggtatacag | 300 |
| agacaagcac | gtttcttctt | ttggtctaata | ttattctaaa | cgaagaagct | gggaactgac | 360 |
| aaaacaggac | agggtgtttt | taatccagtc | tacaaataaa | caagacaatg | cctgagttag | 420 |
| ccctctatat | agatttaggc | ttatgctgac | ctcgttgtaa | aatctgtatt | taactaaaag | 480 |
| ttaataaaaa | tacatatggt | catttttaaaa | taattactga | ttttgcttgg | ctatcccacc | 540 |
| ccttaccccc | aaactcatat | attttttagga | caagattttc | ctgcataacc | acaacctgtc | 600 |
| tcctcccccc | cacccccatc | ata | | | | 623 |

<210> 159

<211> 422

<212> DNA

<213> Homo sapien

<400> 159

| | | | | | | |
|------------|-------------|-------------|------------|-------------|------------|-----|
| aggtaccatc | ttcttcagaa | ctgcatctaa | gaggctgtgc | tggctgggaa | tcatacagct | 60 |
| gtgggcaaca | actgcatcag | ccccaggct | tccctccaga | ccaaaagggtg | attcatggcc | 120 |
| cctggttaat | atcacccctag | gttctcccct | gtcccagttt | taacataata | tttcatagaa | 180 |
| atactagtgc | cataaaaagt | caacattttca | aataataaaa | ttattttata | caaagtgaat | 240 |
| tcataatcat | tcttttataaa | tacagcattg | ttatatatgt | ttgaaacatt | attaaaataa | 300 |
| atattttcta | gagaaaaaat | tttgcttcac | aaaattataa | aacagaagca | tataaaacta | 360 |
| attcatgatt | ggtgcttctt | cagtgtgtct | ctcattctct | cttagttag | acagcatgaa | 420 |
| gt | | | | | | 422 |

<210> 160

<211> 393

<212> DNA

<213> Homo sapien

<220>

<221> misc_feature

<222> (1) ... (393)

<223> n = A, T, C or G

<400> 160

| | | | | | | |
|------------|-------------|------------|------------|------------|------------|-----|
| agctcactct | tttatctgtg | tggctgattt | cattactgtt | tgtgatttgg | agctactcac | 60 |
| tggatggtga | cctcttttca | ctttctctac | tccatgtctg | ggcatgacc | agctttggac | 120 |
| tccttgagcc | cctctctaata | ttaaatttga | tattattaat | tatccaggta | attgtcttcc | 180 |
| gtgtggttgc | ctccttcccc | actccagtat | ccactttcag | caaaacgtct | tgcttcaagt | 240 |
| cccagataga | agagtctttg | acttttcttc | agaggcttat | tttagctaga | atgtttaaag | 300 |
| ctacagatgc | ctatctgctc | atctttccag | ctggattagg | tgttgcttag | atttgctagt | 360 |
| tgctttaagt | attacacagt | ttttgnattt | atg | | | 393 |

<210> 161

<211> 223

<212> DNA

<213> Homo sapien

<400> 161

| | | | | | | |
|-------------|------------|------------|------------|------------|------------|-----|
| accacttaat | tactggcact | gagtatcact | gaatttctta | gttttctagt | ggggaaacat | 60 |
| tattgagaag | ccctccctta | ttttaagtaa | gttgattaaa | tcttatgtga | gttgccagtt | 120 |
| gtaatttttc | aaaggaaaaa | ttttgatggg | gtggaggaat | gaattgccag | ataatctttc | 180 |
| tgggaattccg | agagaattcc | aaagagggtt | tttttttttt | tag | | 223 |

<210> 162

<211> 487

<212> DNA

<213> Homo sapien

<400> 162

| | | | | | | |
|------------|------------|-------------|------------|------------|-------------|-----|
| acaagtctac | attcccacta | acagtgttta | aacgttcctg | cctctgcatt | ctcgtcagca | 60 |
| tttgttactg | tcttttggta | actgtcattc | taacgggggt | aagacaatct | ctcattgtgg | 120 |
| ttttgattct | ctttagaacg | aatattttctc | ctcattcctc | tactcttaat | aatggatttt | 180 |
| ctgaaaaaca | tctattaatt | ttatgcacta | ttcaattcaa | acaacttttt | aaaagttgcc | 240 |
| aaatctgtca | caaaatatta | aacaacaaga | aaaatatcta | aaggtaaact | tgagaggggt | 300 |
| gtaaaacaaa | agactctgag | agcgcactta | gctgtaaaac | aatcattcct | attcctaaat | 360 |
| tgagtgtttt | tggttacatg | ttctaagtgc | cttacaataa | accaggcaat | gtgctttatc | 420 |
| tggagaaagg | gagccctaac | ttcaaagttt | gagttcctcc | aactttttta | atagttaaata | 480 |
| ttcaagt | | | | | | 487 |

<210> 163

<211> 500

<212> DNA

<213> Homo sapien

<400> 163

| | | | | | | |
|------------|------------|-------------|------------|-------------|------------|-----|
| acactggatg | cagccatgca | tggatggttt | ttctttatth | ttcagtgatt | tcctctgaag | 60 |
| cagctgcact | gatacatttg | ggagttgggtg | gcttgacttt | gtccataagg | ggcgtggcca | 120 |
| cttcacatga | tggcgggcct | ttaagagcac | aaagaagttt | aatatggaca | acaacaggaa | 180 |
| aaagcaagaa | gaaaacaagt | agggaaaaac | agctaacctg | gagagaaaga | atttctttaa | 240 |
| cctttatggt | cttcattaaa | aatcttatct | tggactgatt | tgagggattt | ttagaaacat | 300 |
| ggccttatth | tatataagca | ttaccttccc | aggaatctth | gttgatatatt | aatttttgat | 360 |
| aaccatttga | ttacttttaa | aattaagtat | atgtgtgtat | atatacatat | gtatgtttat | 420 |
| atacacacat | gtatctgtat | agttttatat | atacatatat | acacatagac | atacagagaa | 480 |
| ccactactth | gtaatagtgt | | | | | 500 |

<210> 164

<211> 547

<212> DNA

<213> Homo sapien

<400> 164

| | | | | | | |
|-------------|------------|------------|------------|------------|------------|-----|
| actgtaatgg | gtttggccaa | atatcatctt | tgatgacctc | tcctaactca | tcagcacctg | 60 |
| catcagaatg | gtcagtaaag | caggtaaaga | agctctctgg | ttcctcatgc | tgccctcttc | 120 |
| tgctggctth | attctgcgtt | tgactcgaac | gtttcgtcaa | atcctttcca | gatttccatt | 180 |
| tgatttcggt | ggacttcgaa | gatggatcac | cactctcatt | cagatgaaat | tccttggaga | 240 |
| gaactttatt | ttcaaagtaa | ggattttcat | caaaataaaa | atctattctg | taacctgatt | 300 |
| taatatcttc | aaattctgtc | acttcaactc | tggtaaaata | atgcagtgcc | tcctcatctt | 360 |
| cctccccaag | cagtgcagac | acttgtggat | ggttgacaaa | tgttgttacc | caaaaatttg | 420 |
| ggattttggc | gatcaattct | gacctcttct | gaaaaaatgg | ttggcggagt | ttgttatatt | 480 |
| tctgtttctac | tttcaaaaac | tcctcactgg | cttgttcatt | aagtctgtct | atttcattth | 540 |
| gtacctg | | | | | | 547 |

<210> 165

<211> 400

<212> DNA

<213> Homo sapien

<400> 165

| | | | | | | |
|-------------|------------|-------------|-------------|------------|------------|-----|
| acaaaactta | caaagaagtc | aaaagtctta | acactcccat | tctccaggaa | ctcttgtctg | 60 |
| tgtcatctgg | taggagggag | gaatcctggg | tccctcaggt | ccttgtcatg | ttagctttth | 120 |
| gatagcttca | atccactcgg | ctcgtctggc | cttgcctgctg | gcctgaatgt | aatagtgtgt | 180 |
| gtcatcctta | gtaatcaact | tgaagaggtt | tccctggaca | ttccctthta | ccccagtggg | 240 |
| aacgccatta | tcttccagag | cagacacgag | tgaaccacga | agagaaaacc | caccactggg | 300 |
| cctgtttctct | tcttttggag | ggatcatagta | atgcaggaaa | gctggatcct | tccttagaac | 360 |

aaagcgacgc accttccagt ttttctctt gtgccctgct

400

<210> 166

<211> 274

<212> DNA

<213> Homo sapien

<220>

<221> misc_feature

<222> (1)...(274)

<223> n = A,T,C or G

<400> 166

| | | | | | | |
|-------------|------------|------------|-------------|------------|------------|-----|
| ggtaccttca | tataataaag | ttaacaaaaa | taataaaata | ttaaaaaaaa | gagccagctg | 60 |
| gcactgccaa | ccaattccta | tagtagcctt | agaaatccta | atcctgtaga | atttctctt | 120 |
| gtagtcaata | agcaccaccn | tcttcaggag | tatttcagtg | tattgttata | tacaccaagc | 180 |
| aagcctgggtg | atgcagctac | ctgagttctc | ttgggttatgg | gtgaatgtta | tcttcattca | 240 |
| taacttcccn | gctttcatgt | aggtggggat | agag | | | 274 |

<210> 167

<211> 478

<212> DNA

<213> Homo sapien

<220>

<221> misc_feature

<222> (1)...(478)

<223> n = A,T,C or G

<400> 167

| | | | | | | |
|-------------|------------|-------------|------------|------------|------------|-----|
| ctttttaaaa | tccaatatat | tctgccaaaga | atatgccttg | atagttagcc | ctcagcccat | 60 |
| aggtgttttt | tgttttttaa | cagaattata | tatgtctggg | ggtgaaaaaa | cccttgcat | 120 |
| ccaaagggtcc | atactggtta | cttggtttca | ttgccaccac | ttagtggatg | ttoagtttag | 180 |
| aaccattttg | tctgctccct | ctggaagcct | tgcgagagc | ttactttgta | attgttggag | 240 |
| aataactgct | gaatttttag | ctgctttgag | ttgattcgca | ccactgcacc | acaactcaat | 300 |
| atgaaaacta | tttaacttat | ttattatctt | gngaaaagna | tacaatgaaa | attttgntca | 360 |
| tactgnattt | atcaagtatg | atgaaaagca | ataganatat | attcttttat | tatggtaaaa | 420 |
| tatgantgnc | attattaatc | ggccaaatgg | ggagnggatg | ntcttttcca | gnaatata | 478 |

<210> 168

<211> 213

<212> DNA

<213> Homo sapien

<400> 168

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| acaaatgtaa | cagtaatgat | aaattctctt | ttccaaggga | aagagaaacg | ctgcagaatg | 60 |
| gacattaaac | aaggcattat | gccctacaag | caagacataa | aatgtctaag | ggaaacttca | 120 |
| gcataaaaat | gttgaacaca | taatgtgaga | taatttgaat | aaataacaac | tgacattctt | 180 |
| tttttaaaaa | aaaagtataa | aaaatagatg | tgt | | | 213 |

<210> 169

<211> 341

<212> DNA

<213> Homo sapien

<400> 169

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| actggctgcg | aggcgccagt | cgatcaatgt | atgacaggag | ctgagacttg | gccacaccag | 60 |
| gatcccccat | cagacagatg | ttgatgttgc | cccggatttt | catgcctcga | ggagactggt | 120 |

| | | | | | | |
|------------|------------|------------|------------|------------|-------------|-----|
| ccacaccccc | gactagcagg | agcagcagtg | ccttcttcac | atcttcatgc | ccgtatatatt | 180 |
| ctggggcgat | tgaagctgcc | agcttttcgt | agaaatcctc | ctctgcaatt | tgcctcagct | 240 |
| cctccctggg | gagctctcca | gccccagact | catcctcctc | actcttggtc | atcttcacaa | 300 |
| tccgatgggc | ttccaggtag | gtttctgaga | gtaaaccctg | t | | 341 |

<210> 170

<211> 543

<212> DNA

<213> Homo sapien

<220>

<221> misc_feature

<222> (1)...(543)

<223> n = A,T,C or G

<400> 170

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| accaatgac | atgcttccat | tttttttagt | tttaaaccac | caaaccaata | tttttccttt | 60 |
| aaattttaat | cttataatat | agaaatctta | tgtaaataaa | atcttgatca | gtttcaaata | 120 |
| aagagaactg | aagtagaaaa | tagaaatgcc | agtaaacaac | ataatgttta | atttacaact | 180 |
| tacattaggg | gtttggggga | atgctaatta | tatattgaga | atatacatta | gaactcttca | 240 |
| aaatgggctc | ttctaataag | gtcactactg | aacaaaattg | ttccctcttc | tgtaaataag | 300 |
| aatagggtta | aatgactagt | caaataaatt | atcttctctc | tgtaaataaa | attaaatctt | 360 |
| actttctttt | aatgaccaac | cttaggtaaa | acaaaaatat | tgtaatacct | gaaattatcc | 420 |
| tccagctttc | tcacctgaaa | atctattgaa | gtgatccctg | gtcatcctaa | taatgggatg | 480 |
| agggaagttt | ccagcagatt | tcaggctgnt | cttaaagggt | ttggtggnca | ttttctcaat | 540 |
| agt | | | | | | 543 |

<210> 171

<211> 280

<212> DNA

<213> Homo sapien

<400> 171

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| acatactaaa | aatattttaa | atagagaata | ttcctcacag | aggacttttt | tctttaatta | 60 |
| ctactaaaaa | aataattaca | aagtccaaac | aggcagagag | atttagcaca | ctgatcacac | 120 |
| gattctccat | catcctccac | gcttgctctg | aagagggttt | aaaaagtcca | gtttctcggt | 180 |
| gatttcgctg | ctccatttag | ccaagggttg | cctggccact | gattggcaca | agtgggtaat | 240 |
| gcgcttggat | aggatcatgt | tgtgtcttgg | aaatttgggt | | | 280 |

<210> 172

<211> 463

<212> DNA

<213> Homo sapien

<400> 172

| | | | | | | |
|-------------|------------|------------|------------|------------|------------|-----|
| caggtaactat | ttaccctatt | aataagttcg | gtctctgctt | gcaatctttc | cattgctcca | 60 |
| gcataaccagg | gttggcaaga | ataatctact | ggtttgggca | cacatgggca | aggcttgact | 120 |
| gcatacacttg | gaaaaaatcc | aacctctcca | gatgctaaat | ttctgccctg | ccaaaacaga | 180 |
| ctgtgtgcat | ctcctttcag | aagttcaacg | gtatcccogg | cctggagctg | taaagggggg | 240 |
| ccttcatgca | gagctggggg | tggtgttcca | gaatagttcc | taatgacctg | catctttggg | 300 |
| aaacctggat | ccacctgttt | aggagtctct | cgcagtcctt | tggtccgttt | ctctggtagt | 360 |
| ttgagtgtcc | cttgttctga | aagaaatgta | aaaattggca | ttgtcagtg | aaagttattt | 420 |
| tgtttgggta | gcaaccttag | ctttctctgc | agagtggtaa | aac | | 463 |

<210> 173

<211> 165

<212> DNA

<213> Homo sapien

<400> 173
accCAAAGAA ctggtggcct caggccacaa aaaggaaacc caaaagggaa agagaaagtg 60
agaagaaact gaagatggac tctattatgt gaagtagtaa tgttcagaaa ctgattattt 120
ggatcagaaa ccattgaaac tgcttcaaga attgtatctt taagt 165

<210> 174
<211> 532
<212> DNA
<213> Homo sapien

<400> 174
actccatctc tttgactgaa taggtcattg atcctatcaa gggataacaa tgtttttgcc 60
actggatggt gatgttecta tccaaatcca cagcaagctg gtgttgcaat tttccagatt 120
catgcagatc cactgacttc agtgtgttga tactggcttt gaagtattcc atccactggc 180
ggatcgtgga atctcccatc aggtatatga gttttcctct caggcattcc ttcattttga 240
ctgtagccaa actacaggag acaggattcc atgtgtttct ccagacatgc ccactgggga 300
ttgtggatgt cattccaaac ttgcattttct ctttcattgc aactgtttct ttgttgcat 360
tggagacact aattgtattg aattttttcca taatctctac acccacattt gacctttcaa 420
agaggctctt ttcttgtttg ctaagataag aaactttctt gttcttagaa tacatgtgag 480
tgagtgcagc acagggcattg tgttgaggcc tcacacagta gaagccttct tg 532

<210> 175
<211> 374
<212> DNA
<213> Homo sapien

<400> 175
taatcacctg actgagctcc aattaactga ggagaaacgg ggtggaggag agggctgggt 60
gctattcaga cttgataatg agattgatct gtcccatgga gagtgaaagt tcagttccac 120
ttctgcctcc ttctttccat gctgtcctca tgctctttat cctcacttcc tcagtccctt 180
caacactcaa aatctgattt tattttctct tcacacgtat caggggcagt ttctgaagtt 240
gctgagggtg aattttcttc acaaacctct ataaaacatc agcagagaac atataaatac 300
attttgatta gcatacattg caaaatttct cccacaatgt caggggatga aagcagggtg 360
tccccactga gagt 374

<210> 176
<211> 428
<212> DNA
<213> Homo sapien

<400> 176
actgcaactg ccagaacttg gtattgtagc tgctgcccgc tgactagcag ctggactgat 60
tttgaataaa aatgaaagca ttaaagggtt tccctacaaa acatttttct ttaaaatact 120
tttgaaatgg ctataagcag ttgactttca ccttgaggag gcacacact gtgtgagggt 180
cagtgattgt tgaccctccc cagcccctcc tgcttcttta agttatctgt gtgcgtgcgc 240
ttcctctcaa tcttctttgc acgctcattt ctttttctct gaccatgag aaaggaaaac 300
ttactgatga taatttttaa atagtgtaat ttattcattt atagcatgtc aggataaatt 360
aaaagaacat ttgtctggaa atgctgccgg gagcctattg tgtaaataga ggtattttgt 420
aaaataac 428

<210> 177
<211> 318
<212> DNA
<213> Homo sapien

<400> 177
acctgaacga agtcgcgggc aagcatggcg tgggccgtat tgacatcgtg gagaaccgct 60

| | | | | | | |
|-------------|------------|------------|------------|------------|------------|-----|
| tcattggaat | gaagtcccga | ggtatctacg | agaccccagc | aggcaccatc | ctttaccatg | 120 |
| ctcattttaga | catcgaggcc | ttcaccatgg | accgggaagt | gcacaaaatc | aaacaaggcc | 180 |
| tgggcttgaa | atttgctgag | ctggtgtata | ccggtttctg | gcacagccct | gagtgtgaat | 240 |
| ttgtccgcca | ctacatcgcc | aagtcccagg | agcgagtggg | agggaaagtg | caggtgtccg | 300 |
| tcctcagggg | ccaggtgt | | | | | 318 |

<210> 178

<211> 431

<212> DNA

<213> Homo sapien

<400> 178

| | | | | | | |
|------------|------------|-------------|------------|------------|------------|-----|
| acttgaggct | tttttgtttt | aattgagaaa | agactttgca | atTTTTTTTT | aggatgagcc | 60 |
| tctcctagac | ttgacctaga | atattacata | ttcctccagt | aagtaatact | gaagagcaaa | 120 |
| agagaggcag | gattgggggc | acagccgctt | cttcagcatg | gaccaagtgg | gccttgggga | 180 |
| ttgcagcggt | ctcgaagtgg | ctgtaggact | cgaatttaca | gaaagccaca | gaggtgcaac | 240 |
| ttgaggctct | gctagcaagc | caccagttag | gctattgggt | aaccaccttt | ctatacagga | 300 |
| gattggaatc | tactttgtca | tttatccacc | acagtgacaa | aggaaaagtg | gtgccgttat | 360 |
| gcaatccatt | taactcataa | acataattact | ctgagtaact | ggccagccat | tcacgcatc | 420 |
| cttcattggg | t | | | | | 431 |

<210> 179

<211> 323

<212> DNA

<213> Homo sapien

<400> 179

| | | | | | | |
|------------|------------|------------|------------|-------------|------------|-----|
| actgcccact | tttacacaag | ctgcagcaga | actcagttct | actgcagggtg | agagtattgc | 60 |
| accatcatta | acataataag | gacctcagaa | tccaaccttg | ccaaagaatt | caactcctag | 120 |
| gctcagatta | atggaagtgc | tgggcacatg | ccacctcctg | ccattgtcac | agttcagctg | 180 |
| tgttggtccc | gacacagctc | cagttccacc | catgacatct | ggctgaggag | gcttatggga | 240 |
| gctggttctc | atgcacagtt | actgtccctc | tctggagggt | cctttaatgg | ggactgtgca | 300 |
| aagcagtgac | actaactgcc | agt | | | | 323 |

<210> 180

<211> 409

<212> DNA

<213> Homo sapien

<400> 180

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| actgtgttcc | tttgcatgtt | tcttctttta | agaatttagc | tccttctgct | gtttctttta | 60 |
| atgcttcaag | taagccttca | tctgctttta | gtcttctatc | cttacttgag | ggataagttc | 120 |
| aatacctttc | ttggcttcca | caccagaggc | cagggcagcc | gtggtggttg | gtctgagctc | 180 |
| agagctactc | tgaggggtca | catttgcttt | ggcgggtgtg | gcctttcctt | tcttgtcatt | 240 |
| tttggaagtg | tcactgggca | cgtcggctat | gtcactagtt | tcaatgccca | tagctctcat | 300 |
| ttggtctgct | ctcttttctg | taattgagag | aaatttcttt | ggatctgata | aagcatccac | 360 |
| gatattctca | aatccatcag | gcacatatgt | tttaagaaca | atattgcaa | | 409 |

<210> 181

<211> 460

<212> DNA

<213> Homo sapien

<400> 181

| | | | | | | |
|------------|------------|-------------|------------|------------|------------|-----|
| acaaagattg | gtagctttta | tattttttta | aaaatgctat | actaagagaa | aaaacaaaag | 60 |
| accacaacaa | tattccaaat | tataggttga | gagaatgtaa | ctatgaagaa | agtattctaa | 120 |
| ccaactaaaa | aaaatattga | aaccactttt | gattgaagca | aatgaataa | tgctagattt | 180 |
| aaaaacagtg | tgaaatcaca | ctttgggtctg | ttaacatatt | tagctttgct | tttcattcag | 240 |

atgtatacat aaacttattt aaaatgtcat ttaagtgaac cattccaagg cataataaaa 300
aaagaggtag caaatgaaaa ttaaagcatt tatTTtggtta gttcttcaat aatgatgcga 360
gaaactgaat tccatccagt agaagcatct ccttttgggt aatctgaaca aggccaaccc 420
agatagcaac atccctaata cagcaccaat tccttccaaa 460

<210> 182

<211> 232

<212> DNA

<213> Homo sapien

<220>

<221> misc_feature

<222> (1)...(232)

<223> n = A,T,C or G

<400> 182

actgacagat taatggcttg cctagagctg tgcaagaaac agcctgccag nctgtcattg 60
nnagggaacca gggcaaaacc aagagctggt cttcccagaa gagccctgca aacacattgg 120
ttcgtgcttc cctttacttc ttctggtcag ataccatgaa tgccagtcac cagtaaattct 180
taatacactt ttgctttatt ctcacatgcc attcaccaga ttatttgatg gt 232

<210> 183

<211> 383

<212> DNA

<213> Homo sapien

<400> 183

atgttattta aaagatgaaa tttcatgggt caaatgtatt tttctcccat aaaaatattt 60
tctcttccat ttāaatatat acctaattct tgagaaatct tgcacaaatg gcattttatt 120
aaagaaaatc taattfaca agctttgtaa attttgagaa aaacattcat agatcataaa 180
caaaaatttc aatatgcaat attcaaatTT acaagaaaat aagcacaaac ttttagacag 240
tgcagttatt gctgcactcc ttttaattcct tatccagagc ccaaaaaatg taggcaaacc 300
ctaaaaatgt agcagaagca tttccgcaca ctgggtgtcca gaattctagt tgtgcagaaa 360
tgtttccact agatttatag agt 383

<210> 184

<211> 444

<212> DNA

<213> Homo sapien

<400> 184

acagacacaa acatataaat atatgtatgc acatatttgt catacatTTT caataaatga 60
tatcttttatt attgttttaat gacctttttt ctcttggtgaa ttttgacata aagtatatTT 120
tataaaataa gagagttggt gacttacgat gtattttgta taatacaatt ttgatctctt 180
ctgctctcat ttggttgatg tttgcctaaa atgtcttctt ccacttgcca ctttcaggct 240
gatttcacta ctagatctca agtgactctt gaagagaggc aagttggatc ttggtatata 300
aaattttata taatccctct attcaatgta tgtgtattga ttggcaagtc tattttttaa 360
atatttatTT tctgaagaca aagattactg ttatttttatt gtttaaatgat tcttgtaggt 420
ctgtttctca ttctatcttc cttt 444

<210> 185

<211> 289

<212> DNA

<213> Homo sapien

<400> 185

acttgtagaca ggcagacgtg attgcagcca cgaacacgat gaactcactg aagtccacct 60
gggcatctcc attggcgtcc aggtccttga gcaatttatc cacggcatcc ctgtcttttc 120

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| cactctgcag | gaagcctggt | agctccttct | ccatcagcac | cttgagctcc | cccttggtca | 180 |
| gggtctgcgt | gctgccctcg | ctgcccgaat | atcgggaaaa | gacgtctatg | atcatgccca | 240 |
| tgactgtctc | tagttccgtc | atggtgctag | attcagaccc | accttcctc | | 289 |

<210> 186

<211> 407

<212> DNA

<213> Homo sapien

<400> 186

| | | | | | | |
|------------|------------|------------|------------|-------------|-------------|-----|
| acagacaaaa | tgctcaggat | gccatgattg | ccctagagca | tggatcacct | tcccagcaat | 60 |
| cggtttcttg | caggatgcac | aatggccctt | gggcactgtg | gcaatgccaa | ggtcctgcaa | 120 |
| ttcctgctcc | agacccccaa | gcattgagtc | cagggaggcc | ttgtgatcct | gcttgtctgg | 180 |
| taagtgtctc | ttgccagcat | ctgctctcac | tgcaaccttg | gcctgcatct | cagtcagggtg | 240 |
| agccatgagc | tcatccaact | gagcagctgc | tgacgtttta | gaagggtggtg | gtgattcctt | 300 |
| tggtctttgg | gcttcactgt | agacattgag | ctcctggata | ttggtagtat | acacgagctg | 360 |
| cgccggcaag | ggacttgtgt | tatcctgaat | agaaaggatc | tccgaag | | 407 |

<210> 187

<211> 441

<212> DNA

<213> Homo sapien

<400> 187

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| actgcaagac | ccatcttccc | tccagttaat | acactcccag | gatgggctgc | agagggggag | 60 |
| actctgagag | aagctggagg | cccacaaaag | tccactgacc | ctctttctgt | cccagaaatg | 120 |
| aataaaggac | ccagttgtgc | tttccttcca | aaatcctcaa | caaagttggt | tgtgctccaa | 180 |
| gaaaatgtgg | gaataaaaaa | atcatgtccc | aggtcattct | tgtgtgtgtg | cggggggagg | 240 |
| ggatgggagg | aaaaggcatg | tattaataga | tactgctgct | ataaaatgac | ataaatcata | 300 |
| gcccttgatc | tgtttctgta | aacaatgcc | gcttcttcag | gttattggca | actacccta | 360 |
| atatacctag | cccagatcct | ttcataaagt | caagtgttat | atttccaaaa | taatcctatg | 420 |
| aatcatgaa | ggttgtgaag | g | | | | 441 |

<210> 188

<211> 323

<212> DNA

<213> Homo sapien

<220>

<221> misc_feature

<222> (1)...(323)

<223> n = A,T,C or G

<400> 188

| | | | | | | |
|------------|-------------|------------|-------------|------------|------------|-----|
| acttagaaaa | cagtcctctgt | ccatcagcca | gaaaagggtga | ccatcacccc | taaagtaatt | 60 |
| tccaaacttt | agttcagtgg | gaaagatatg | ctggtagtgc | atattcagng | ntgattttca | 120 |
| gtgctagtaa | ccacttttaa | tgccagaaat | atgtaacaat | gataatgtaa | cgtcaaagtg | 180 |
| gttactaaag | attatagcct | taactttttt | atgnaaaaga | taaaatccat | tcctcctccc | 240 |
| agttagcaag | catggcttgc | atttctcaaa | aatgagaact | tccatggcag | ccaagaaaac | 300 |
| gtcttctcag | aggaactttc | gtt | | | | 323 |

<210> 189

<211> 225

<212> DNA

<213> Homo sapien

<400> 189

| | | | | | | |
|------------|------------|------------|------------|------------|------------|----|
| caggtactcc | ctgatctttt | cctcagtggc | ttcaggattc | agacccccaa | cgaagatttt | 60 |
|------------|------------|------------|------------|------------|------------|----|

cttcaccggg tccttcttca tagccatggc ctttttaggg tcaatgacac ggccatccag 120
cctgtgctcc ttctgggtcta ggaccttctc cacactggct gcattcttga acaggataaa 180
cccaaaccct cttgaccgtc cagtgttggg atccattttt attgt 225

<210> 190
<211> 501
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(501)
<223> n = A,T,C or G

<400> 190
acagctgaag ttngataaca aagaaatata tataagacaa aaatagacaa nagttaacaa 60
taaaaacaca actatctggt gacataacat atggaaactt tttgtcagaa agctacatct 120
tcttaatctg attgtccaaa tcattaaaat atggatgatt cattgccatt ttgccagaaa 180
ttcgtttggc tggatcatac attaacattt tcnagagcaa atccaagcca ttttcatcca 240
agtttttgac atgggatgct aggttctctg gnttccattt gggaaatgta ttcttatagn 300
cctgtaaaaga ttccacttct ggccacactt cattattggg agtgcccaaa gctctgaaaa 360
tcctgaagag ttgatcaatt tctgaatccc catggaaaag tggtttctta gttgctagtt 420
cagcaaatat ggtgcctata ctccaaatgt caactggagt tgagtaacga gctgacccca 480
gcaatacttc tggagatctg t 501

<210> 191
<211> 436
<212> DNA
<213> Homo sapien

<400> 191
acagtgcatt gtgctgtcac ttggaaagcc tttcaatggt gtcttcagat tgttgtgatg 60
aatatgaaac atgcagaccc tcctttataa agaaaaagac cttaaaactt gaatatgaga 120
taattttaca ttttaaaagt ttatttgatt ttcatattat tcactttcaa agccctttca 180
aatagaaaag gtatgaactt ttggggggat aatttatgta tcgtaaactt attagaacaa 240
aatattcctg atgtataatg agttgtttta ttataacaac tttttcaatg gtagtttgca 300
ctattcttta ttatgctaca ggtttattta ttatgaaaca aaggaatatg tattttatgt 360
attttacat gcataggtta actctttgcc acagatttat tggttcttga tacacctaaa 420
ataaaaaaaa atgtgt 436

<210> 192
<211> 319
<212> DNA
<213> Homo sapien

<400> 192
ccagcgacag actttgcaaa catgcagatg gttctcacat gtcttccttg tctcattttc 60
agggcacgtg tcctagggtc tttcgattac gtctctcaag gcaaggtttc cagatctctc 120
tgtatcctta cgcttccctt ttggatgcac ctttaattta aaatacctct ttttctcatt 180
aattagatca cttcaagtta aatacaaaac atggcaagat ggatttaaatt ttagagggat 240
ataagtatac ataagagaag accaatctct acttttataa atgcagttta ttaacaataa 300
agtaaaatat agtgaaggt 319

<210> 193
<211> 586
<212> DNA
<213> Homo sapien

<400> 193

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| acaagaggcc | atgtgtcttg | cctttttctg | acatgtgcat | actataaaat | cacaggtagc | 60 |
| caacatttag | tatcagtaaa | aaacaactac | gtttgttcac | ctgtttggca | tagggagaaa | 120 |
| acaatgtatc | tcatagcatt | aaatgataca | gccttaacac | atatgatgct | catatttgca | 180 |
| aagttcccaa | atgttgagaa | gttctagtga | aaagtcatac | tattgtgcaa | agatgaaaat | 240 |
| ttggggccaa | tgtctgtatt | caaaataacc | aaaatatatt | ttaaagcaaa | atatatcctg | 300 |
| atactactat | agattctagg | aattgtccta | aaagagtaaa | gtgttggttc | ctttctgaac | 360 |
| atgaataaca | tcaaaggaag | aacccagttc | ttaagactta | agtaggaaat | ttatagaaat | 420 |
| ttgatttata | ccagtagtaa | taacattcat | aaggaaaaac | tattaggtaa | caattttctc | 480 |
| caagaagagg | atcagattac | ttaaaattgt | tggagaattc | tggttgtttg | cgcaataatc | 540 |
| atagtgtatt | acattgcttt | tcttctttca | gagcaataag | aaagtt | | 586 |

<210> 194

<211> 214

<212> DNA

<213> Homo sapien

<400> 194

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| acatttttat | aactggaatg | tttatgtgta | gtgaagctct | gagaggactt | tgcattagat | 60 |
| ctcagcagca | taatcagaag | gttgctcttt | gtctcagcaa | tttttaagct | aatagtagca | 120 |
| gaaattgcag | tggaaataga | ctgctttgcc | acaacattca | gaaaatcatt | tatcttttta | 180 |
| ttgcagttct | tgtcaccaaa | caatacattt | tagt | | | 214 |

<210> 195

<211> 325

<212> DNA

<213> Homo sapien

<400> 195

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| actgtacata | tttgcaatca | cattgtgcat | agattcttaa | tggtagatat | gatttctttt | 60 |
| gtcaggctac | aacaatgaac | tgcagattcc | ttgtttgtaa | tgtaaatgat | tgaatacatt | 120 |
| ttgttaatat | gtttttattc | ctatgttttg | ctattaaaaa | ttttataaca | tttccaagac | 180 |
| aaaaattcca | agtttatgct | ttgaagaatt | tatgtaatta | aaatttcact | aaactaatct | 240 |
| ttttagttta | ggaattattt | gggttttgac | actggaagtt | gcgccaaata | agcatcagaa | 300 |
| ataggagatg | cttaacattg | ctata | | | | 325 |

<210> 196

<211> 382

<212> DNA

<213> Homo sapien

<400> 196

| | | | | | | |
|-------------|------------|------------|------------|------------|------------|-----|
| actccttccc | agttttttct | ttatactgag | ccttcaggga | cagtaagcat | tctacagctt | 60 |
| catttatattt | agccttaggg | gatttttcag | cttttagctt | acgaaccacc | tccccttggt | 120 |
| cagcaacttc | atcatacaga | gatttacttt | ccagaatact | tgctgaggaa | ttagaagaaa | 180 |
| tattctgtcc | tatttcagca | ggagggtttc | caggtttata | ttcctggcca | gttttctcct | 240 |
| tatattcagc | tttcaaagac | aaaagctggt | ttacagctgc | atctacatct | tcctttggtg | 300 |
| ctttcttggc | ttttaattca | cgaaccacat | ctccttgaa | agccactcta | ttgtaaagga | 360 |
| ccaaggaatc | ctcagatgta | gt | | | | 382 |

<210> 197

<211> 648

<212> DNA

<213> Homo sapien

<220>

<221> misc_feature

<222> (1)...(648)

<223> n = A,T,C or G

<400> 197

| | | | | | | |
|-------------|------------|------------|------------|------------|-------------|-----|
| acatccacat | gttcctccaa | atgacgtttg | gggtcctgct | tgccaacatt | ctttattgcc | 60 |
| agctgttcag | gtgtcatctt | atcttcttct | tctacagcct | tattgtaatt | cttggctaata | 120 |
| tccaacatct | cttttaccac | tgattcattg | tgtttacaat | gttcaactga | gtcctgaagt | 180 |
| gtcaaacctt | ccatccaact | cttcttatgc | aaatttagca | acatcttctg | ttccagttca | 240 |
| tttttccgat | agttaatagt | aatggagtaa | taatgtctgt | ttagtccatg | aattaatgcc | 300 |
| tggaatagatg | gcttggttaa | gtgaccaga | ttcgaagtgt | tttgtcttgg | ttcatgtcct | 360 |
| aagaccatca | tattagcatt | gatcaatctg | aaggcatcaa | taacaacctt | tccttttaca | 420 |
| ctctgaatgg | gatccacaac | cactgccaca | gctctctccg | acaaggcttc | aaagctctgc | 480 |
| tgagtgttga | tatccacacc | agaaagccaa | caaccaaagc | cagggtgact | gtgataccaa | 540 |
| ccaacaacca | tctccggcct | tcctgtctgc | ttcaacatat | ccaacatttt | aacttggaac | 600 |
| actggatcaa | ctgccttcac | actgacacct | ggtnctgatg | nggcataag | | 648 |

<210> 198

<211> 546

<212> DNA

<213> Homo sapien

<220>

<221> misc_feature

<222> (1)...(546)

<223> n = A,T,C or G

<400> 198

| | | | | | | |
|------------|------------|------------|-------------|------------|------------|-----|
| acaatacagc | accactactg | agaagggtc | gaggttttgc | aatccaaggt | tctgacttaa | 60 |
| agcaaaaata | cacggcatag | attgcaacag | caaagaagtg | tccaattaaa | actagagggt | 120 |
| taggagacaa | tacagaaagc | agcccaacag | gacccgcaac | acattcgcca | ccaagtttga | 180 |
| aataaagaaa | acagggtttt | cttagttgat | gcagggaatc | atctgtggca | gaaaataatt | 240 |
| cataaagagc | ctgagcaagg | atattcacga | caaagggaatg | agatgttttt | cttgcccagt | 300 |
| aaaatgattt | tttggcctcg | aaaatagctg | catcatcata | aagggtcagg | atacccttta | 360 |
| gcagttttct | ccatagtttt | atatctttaa | aagcaacagt | cattcctcca | ccagtaagtg | 420 |
| gatgcctcat | attatatgcg | tctcccaaaa | gaagaacacc | tcgtttcttc | actgatgaag | 480 |
| gaggaaggaa | gcttgctgca | tggacctcag | atgagaattg | cagtggttct | aagaatggtc | 540 |
| ntttca | | | | | | 546 |

<210> 199

<211> 275

<212> DNA

<213> Homo sapien

<400> 199

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| actatgtgta | actttggcaa | caggttgcag | tcagccaggg | tgagctcggt | gccatccaaa | 60 |
| aacttcctct | gagagacacc | ttcatcttca | gcactgggtt | catccacttc | ttctgggagg | 120 |
| ggggatgtta | agtaattgtc | taaaaccttc | agggctttca | ggagtccctt | ctccagattg | 180 |
| tcattgagtg | ctgggtttga | attcttgatg | taggcagaaa | atttggcaaa | tatgtccagc | 240 |
| ccagctgtgt | tggactcagg | gttcagagct | gccag | | | 275 |

<210> 200

<211> 423

<212> DNA

<213> Homo sapien

<220>

<221> misc_feature

<222> (1)...(423)

<223> n = A,T,C or G

<400> 200

| | | | | | | |
|------------|------------|------------|------------|------------|-------------|-----|
| cctgagaaat | tctnaaaagt | acgatgataa | ggttgcaaaa | atgaagaagc | tcatcatact | 60 |
| aaaactagga | aacatacnga | tccataacan | gacatgcnaa | gcaaagttcc | caaagtcaca | 120 |
| gacaagaaga | gaatctcaaa | tgctggaaaa | tacataatta | tggttgcatg | atntaaccag | 180 |
| tgactctttc | aacataaacc | ttgcaggcca | gaaggaaatt | gcgtgctata | gttgagggtgc | 240 |
| caagcgaaaa | atagcttcta | tgtaagaata | acataaccag | caaaactgtg | ctacaaaaat | 300 |
| gaagaaaaag | caaagacctc | taaagataac | caaacgtgga | aaaattatat | caacactaca | 360 |
| tgtgccatac | aaaaaatgct | gagaagagtc | ctcctattaa | aactatatga | tgctaaaaaa | 420 |
| caa | | | | | | 423 |

<210> 201

<211> 560

<212> DNA

<213> Homo sapien

<220>

<221> misc_feature

<222> (1)...(560)

<223> n = A,T,C or G

<400> 201

| | | | | | | |
|------------|------------|-------------|-------------|-------------|------------|-----|
| acaatcgagt | attttagaaa | ttacatgaaa | catgaaacag | tttttgcaat | tttttttaaa | 60 |
| ctgggcatct | ggtttctaaa | aatttatattg | aaacaatcta | gaattttctt | ggtgcaaagt | 120 |
| gtatcatgtg | gaatatectc | atatttttac | catattttta | gaactttaag | acgattaatt | 180 |
| gtaaataatt | tatttgattg | gtgcagttct | aatccctaaa | tcataatctt | aaaatcagga | 240 |
| atgtgtggag | aacagagcca | tgtcatatca | ctttgctctt | accattcctt | ttgatcagcc | 300 |
| tcaattcagc | ctcattgtgt | agtatgtttt | ttctttctat | gaaaaacaac | agaaagcatt | 360 |
| tcattttatt | tgcctatgtt | caaatatgtt | taataatgac | caaagtgcac | tctgagtttt | 420 |
| ttcaaggaat | gtaatactgg | agctttaaga | acatacttag | tttctcatgt | gaaaacttan | 480 |
| gctttgtctg | angttttcct | tcctctattg | nctaattggtg | aggtgggtttt | aggaattatg | 540 |
| ttttataact | tttcaatata | | | | | 560 |

<210> 202

<211> 366

<212> DNA

<213> Homo sapien

<400> 202

| | | | | | | |
|------------|------------|------------|-------------|-------------|------------|-----|
| acgagcccca | cagagcagga | agccgatgtg | actgcatcat | atattttaaca | atgacaagat | 60 |
| gttccggcgt | ttatttctgc | gttgggtttt | cccttgccct | atgggctgaa | gtgttctcta | 120 |
| gaatccagca | ggtcacactg | ggggcttcag | gtgacgattt | agctgtggct | ccctcctcct | 180 |
| gtcctccccc | gcacccctc | ccttctggga | aacaagaaga | gtaaacagga | aacctacttt | 240 |
| ttatgtgcta | tgcaaaatag | acatctttta | catagtccctg | ttactatggt | aaccttttgc | 300 |
| tttctgaatt | ggaaggga | aaaaatgtag | cgacagcatt | ttaaggttct | cagacctcca | 360 |
| gtgagt | | | | | | 366 |

<210> 203

<211> 409

<212> DNA

<213> Homo sapien

<400> 203

| | | | | | | |
|-------------|------------|------------|------------|------------|------------|-----|
| cgagggtactg | aagaacccca | tcatgtgaga | gatcgctcaa | agtcattaac | acaaagcagt | 60 |
| gaaaatcatc | cagcaaagca | gtgctattat | gagtgtgggc | tatggaaaga | cagcttttcc | 120 |
| tacactgata | aagaaaaaaa | aatgaggaaa | ttatttcatc | cccttgtagc | atctgtgact | 180 |
| ttttggattt | aataatcttg | ctgtttttcc | tctttatgac | aaagaatata | attgggagga | 240 |
| tgaagtgtct | taaaaattgt | agagaccagc | tcactggaat | gtttttccat | ccctgtattc | 300 |

atggcttgac tttgtgactg ctctacactg catgtctgac attgcagagt gagctatgtt 360
gaggtaaact gggttggttg attatttttg aatcagcctg gtctctccc 409

<210> 204
<211> 440
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(440)
<223> n = A,T,C or G

<400> 204
acacacatcc tgatctagct atgtttatgt gtgttggggg gatggatgga caagaggtat 60
agttcaaagt agatcatttt tgtgaaatgg ctttgtaaac tgtaacatgc cctataaata 120
tgagattagc tttaataactg gccctgactc tccagtgtgg ctttgtgtgt ttgtctaaac 180
acttagttaa tatctgtcag tgggccattg cacaaggaac tgacacaatg gtatcctgtg 240
cctctgttgt tgttggttgt gttttttttg cagttctaaa agcttagtta attgccttca 300
ttagcttaat atataccacg tgaaaagcat agaaaagcag aactcaaaac tcanagaata 360
aaggacagaa cataactaac tactgatgtg caccttagtt acctgatgca gggaattgaa 420
gcatataagc ttcacttagt 440

<210> 205
<211> 474
<212> DNA
<213> Homo sapien

<400> 205
acttgtccca tgctaggtaa caggaaaata atagtgtattg ataagacata gtccctgtcc 60
tcaaagagtt aacagtctag caaggcagga actttgagaa aagaccaatg tgttcaaagg 120
aaaactcaca acctgggtct cccttctcag atggcacatt caagaaactg ttgcttatgc 180
ccctgggagc cagagcctta cttaagtctt accaagtcaa atatctatca gcttcagatg 240
atttgagcct ggtaaagtct tagcaataga tttgctgcct catgttccca tgaaaacctg 300
ataagagaga gccctttcaa ctcaggcata cgggggggtt aaggataaca tgtttagtga 360
ccatgtggac attcagcaca ggtgagcttc tcaagtgaga gccatgtgtc cccaaaagaa 420
aggagggttt atccataaga ctttgctctc cctttcaaca ctgtggtggg aagt 474

<210> 206
<211> 344
<212> DNA
<213> Homo sapien

<400> 206
accgtccttc ttggggcaga tgtctgagat aaactgttcc acgccccag ccaaaccaca 60
gcagttcaac gcatagtga tggctttcag cgtttcccgc tgggggtcat ccttggtttt 120
cagcttggtg taggtgtcct tgtaaaactc ctggacttcc ttaatcacct catccttgtg 180
ggaatatccc cagatggccg cagctatttc aatggcgaat atcaccaaga ggaagccgaa 240
gaacagtccc agcatgcact gggactcctg cacagccccg cagcagccca ggaagcccac 300
cagcatcatg agggcgccgg ctccgatcag aatatagact cctg 344

<210> 207
<211> 441
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature

<222> (1)...(441)

<223> n = A,T,C or G

<400> 207

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| acctcaattt | ttcccccaat | ttctggctac | tactaaaagc | cagaaagaaç | agaacagtgg | 60 |
| cctcaggaga | tctgagtttg | aatccttgct | ctctaggatg | caggtggctt | gaagcagaat | 120 |
| gccacacctg | caagttgatt | agaactgcct | ttcttcccag | gcttgacata | ggtattaagt | 180 |
| caaaattaca | tgaaaccag | tggtaaaaaa | gcctctgaaa | gctgtaacac | cctcagtaat | 240 |
| aacaaaagg | atttttat | cacagctaaa | gggaaaatag | gtggagaagt | taaaaaataa | 300 |
| tgtctgatcc | tgttcctaag | ttccaaacta | tagccaacac | tctgatgctg | ctctttttct | 360 |
| tgtaggacca | accgtcccag | tttgcctggg | actttctcat | ttttacagag | tcccaaattc | 420 |
| tangaaactg | gagcaactgg | t | | | | 441 |

<210> 208

<211> 365

<212> DNA

<213> Homo sapien

<400> 208

| | | | | | | |
|-------------|------------|-------------|------------|------------|------------|-----|
| ctgggtgccag | tgccagtgtc | tgagccagtg | ccagagccgg | aacctgagcc | agaacctgag | 60 |
| cctgttaaag | aagaaaaact | ttcgccctgag | cctatttttg | ttgatactgc | ctctccaagc | 120 |
| ccaatggaaa | catctggatg | tgcccctgca | gaagaagacc | tgtgtcaggc | tttctctgat | 180 |
| gtaattcttg | cagtaaatga | tgtggatgca | gaagatggag | ctgatccaaa | cctttgtagt | 240 |
| gaatatgtga | aagatattta | tgcttatctg | agacaacttg | aggaagagca | agcagtcaga | 300 |
| ccaaaatacc | tactgggtcg | ggaagtcact | ggaaacatga | gagccatcct | aattgactgg | 360 |
| ctagt | | | | | | 365 |

<210> 209

<211> 191

<212> DNA

<213> Homo sapien

<400> 209

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| cgaggtacag | aatataaagg | agactgttga | attcatacca | tataaaactt | gttaggtttt | 60 |
| taaacatagc | aatcaaggct | acaaaaacaa | acctgtgttg | tttttgata | gattgtaggt | 120 |
| ttatttttgg | attcatata | catgactgaa | ctgtgtgcaa | ggcaatagtt | agccttgatt | 180 |
| ttagcccaga | g | | | | | 191 |

<210> 210

<211> 373

<212> DNA

<213> Homo sapien

<400> 210

| | | | | | | |
|------------|-------------|-------------|------------|------------|------------|-----|
| acttaattgt | atatttcatt | taaatagtcc | ttctcagggg | tttaataatt | tagaatcaat | 60 |
| agttcccttc | aaaacataat | aaaatattta | cactttataa | aatattaacc | cgattaacaa | 120 |
| tacagccgtg | ttgtttataa | gagtgttaact | gaagtcctgc | aaatcatgct | gttgacacaa | 180 |
| gcctgtgagg | ttagecgaagt | gatccttagc | aaaatgtaaa | tgaagatctt | cagacagtgg | 240 |
| tgtttataaa | atagctcatt | aatgacttag | gattgaatcg | ctccaacat | tcgcatcatc | 300 |
| agatataata | atagtgacga | atcagacagg | aaagatcctg | gctaaacat | ttgcattttt | 360 |
| ttccagaagt | acc | | | | | 373 |

<210> 211

<211> 336

<212> DNA

<213> Homo sapien

<400> 211

actgtaatct ttcttcatca aaatatgcaa aacagcatca tggattgtta agaaaaatat 60
tgagcttttc acttcacccat caaaaaattc ataccgggta agcttctcaa tgaagtcac 120
atcagttcca acgatataca catctacctt gatcctgata aattcttgca aaatcgattt 180
aaggcccctc actgaagaaa catcaagaaa ggacactgct gaaaagtcga gaatgaggct 240
gtggaggctg attttgggga cctcaatgtt gagaggaaga tcatcattcc agtcaatgtg 300
gaaaggcagg tctgtggtat tgattgctgg tccagt 336

<210> 212

<211> 434

<212> DNA

<213> Homo sapien

<400> 212

accaccagca attttaagga aatcttcacc tgttgctttg taaacctcaa tataccgggt 60
ccccatgtga tgtttgtgcc tctgtagtgc taggtctcgg tgctcctcac ttacaaacct 120
aaccagagct tctccgttcc ttcgaccctg agcattcaga caaagtgtg cactccctt 180
ggcaatattg agtcccttga agaattctgc aatatcttga tctgaagact gccatggtaa 240
acctcgtgcc ctgactacgg tgttatcatc aataagttcc atcttgctgc aagttccact 300
ttcaaacttg taattcactc tctctggatc tgaaaacctg tgattataag gctctgaaat 360
cattgctaaa attatattcc ccatacttcc aacttgagag gctccatata gagagactga 420
actactcttc tcaa 434

<210> 213

<211> 515

<212> DNA

<213> Homo sapiens

<400> 213

actacacgac acgtactctt gaatacaagt ttctgatacc actgcactgt ctgagaattt 60
ccaaaacttt aatgaactaa ctgacagctt catgaaactg tccaccaaga tcaagcagag 120
aaaataatta atttcatggg actaaatgaa ctaatgagga taatattttc ataatttttt 180
atttgaaatt ttgctgattc tttaaattgc ttgtttccca gatttcagga aacttttttt 240
cttttaagct atccacagct tacagcaatt tgataaaata tacttttgtg aacaaaaatt 300
gagacattta cattttctcc ctatgtggtc gctccagact tgggaaacta ttcataaata 360
tttatattgt atggtaatat agttattgca caagttcaat aaaaatctgc tctttgtatg 420
acagaatata ttgaaaaca ttggttatat taccaagact ttgactagaa tgcgtatttt 480
gaggatataa acccataggt aataaaccca caggt 515

<210> 214

<211> 353

<212> DNA

<213> Homo sapiens

<400> 214

acaagactca agtaaataga aaggcagctt tcaatcacaa atcagttttt cagatttttac 60
tgtggaagca tatttaaatgc acacatttga atgttacaca taaataattt taacgatgga 120
gtccaagttc tggattttac attagatctg catatataag acacttgtgg tcaaatttca 180
agattggtta agccagtttc aagctgctta tattttgagt acctgccgg gcggcgctaa 240
ggcggaattc tgcagatata catcacactg ggccggccgct cgagcatgca tctagagggc 300
ccaattcgcc ctatagttag tctgattaca attcactggc cgtcgtttta caa 353

<210> 215

<211> 699

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature
<222> (1)...(699)
<223> n=A,T,C or G

<400> 215

| | | | | | | |
|------------|-------------|------------|-------------|------------|------------|-----|
| acacttgaaa | ccaaatttct | aaaacttggt | tttcttaaaa | aatagttggt | gtaacattaa | 60 |
| accataacct | aatcagtggt | ttcactatgc | ttccacacta | gccagtcttc | tcacacttct | 120 |
| tctggtttca | agtctcaagg | cctgacagac | agaagggcgt | ggagattttt | tttctttaca | 180 |
| attcagtcct | cagcaacttg | agagctttct | tcagtgtgtc | aagcaacaga | gctgtatctg | 240 |
| cagggttcgt | agcatagaga | cgatttgaat | atcttccagt | gatatcggct | ctaactgtca | 300 |
| gagatgggtc | aacaaacata | atcctgggga | catactggcc | atcaggagaa | aggtgtttgt | 360 |
| cagttgtttc | ataaaccaga | ttgaggagga | caaactgctc | tgccaatttc | tggatttctt | 420 |
| tattttcagc | aaacactttc | tttaaagctt | gactgtgtgg | gcactcatcc | aagtgatgaa | 480 |
| taatcatcaa | gggtttgttg | cttgtcttgg | atttatatag | agcttcttca | tatgtctgag | 540 |
| tccagatgag | ttgggtcaccc | caacctctgg | agaggggtctg | gggcagtttg | ggtcgagagt | 600 |
| cctttgtgtc | ctttttggct | ccaggtttga | ctgtggtatc | tctggccaga | gtgtaggaga | 660 |
| nggccacaag | gagcaagaat | gctgacactg | gaattttct | | | 699 |

<210> 216

<211> 691

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<222> (1)...(691)

<223> n=A,T,C or G

<400> 216

| | | | | | | |
|-------------|------------|------------|-------------|-------------|------------|-----|
| ncgaggtaca | ggtttacta | ttacaaatat | atgatgttaa | actaacaac | tcattgacct | 60 |
| caaagatgtc | ttcgteccac | gcacacacat | ttgtaatttg | tgtccatttg | ctatttccct | 120 |
| tcttctataa | tcttcaaatt | atatagttat | gcattgagtt | ccctatgcat | ctcaccctac | 180 |
| tccttttatct | cagccttctc | atactttgcc | attctcttct | ttctggaaat | aactagcaca | 240 |
| acaattccag | caacaactgc | tatcaccaca | accacaataa | cagcaataac | accagctttt | 300 |
| agaccctgca | ttgagaattc | aggtgctttt | tcattcaacat | aataaattaa | agtttgacca | 360 |
| ggatccagat | ccagttgttc | cccatttact | gtcagggtcca | ttttcttaga | atgaaacaag | 420 |
| gattcacctt | taacatcttt | ttcaaaataa | taagccacat | cagctatgtc | cacatcattc | 480 |
| tgagtttttt | gagaagaatt | ttgaaccaga | tcaatagtga | taacattatt | ctcatacaaa | 540 |
| atactcgtga | taaatttttg | atccagttga | taacgcgttg | tgatctcctt | ctgaagtgca | 600 |
| gtccgcaaac | ttttactatc | ataaggggtt | tctcttgctt | tgnggttttag | ttcaatggat | 660 |
| gatccagtag | ggtctcactc | gctcagagca | a | | | 691 |

<210> 217

<211> 497

<212> DNA

<213> Homo sapiens

<400> 217

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| ctgtgctcct | ggatggtttt | accacaagtc | caattgctat | ggttacttca | ggaagctgag | 60 |
| gaactgggtc | gatgccgagc | tcgagtgtca | gtcttacgga | aacggagccc | acctggcatc | 120 |
| tatcctgagt | ttaaaggaag | ccagcaccat | agcagagtac | ataagtggct | atcagagaag | 180 |
| ccagccgata | tggattggcc | tgcacgaccc | acagaagagg | cagcagtggc | agtggattga | 240 |
| tggggccatg | tatctgtaca | gatcctggtc | tggcaagtcc | atgggtggga | acaagcactg | 300 |
| tgctgagatg | agctccaata | acaacttttt | aacttgagac | agcaacgaat | gcaacaagcg | 360 |
| ccaacacttc | ctgtgcaagt | accgaccata | gagcaagaat | caagattctg | ctaactcctg | 420 |
| cacagccccg | tcctcttcct | ttctgctagc | ctggctaaat | ctgctcatta | tttcagaggg | 480 |
| gaaacctagc | aaactaa | | | | | 497 |

<210> 218
<211> 603
<212> DNA
<213> Homo sapiens

<400> 218
acaaaggcga aagagtggat ggcaaccgtc aaattgtagg atatgcaata ggaactcaac 60
aagctacccc agggcccgcg tacagtgggc gagagataat ataccccaat gcatccctgc 120
tgatccagaa cgtcacccag aatgacacag gattctacac cctacacgtc ataaagtcag 180
atcttgtgaa tgaagaagca actggccagt tccgggtata cccggagctg cccaagccct 240
ccatctccag caacaactcc aaaccctggt aggacaagga tgctgtggcc ttcacctgtg 300
aacctgagac tcaggacgca acctacctgt ggtgggtaaa caatcagagc ctcccgggtca 360
gtcccaggct gcagctgtcc aatggcaaca ggaccctcac tctattcaat gtcacaagaa 420
atgacacagc aagctacaaa tgtgaaaccc agaaccctagt gagtgccagg cgcagtgtatt 480
cagtcacccct gaatgtcctc tatggcccgg atgccccac catttcccct ctaaacacat 540
cttacagatc aggggaaaat ctgaacctct cctgccacgc agcctctaac ccacctgcac 600
agt 603

<210> 219
<211> 409
<212> DNA
<213> Homo sapiens

<400> 219
ctgagagacc aggagaagtt ccagatgcag agactgtgat gctcttgact atggaattat 60
tgccggccagt agccaagtta gagacaaaac aggcgtaggt cccgttatta tttggcgtga 120
ttttggcgat aaagagaact tgtgtgtgtt gctgcggtat cccattgata cgccaagaat 180
actgcgggga tgggttagag gccgagtggc aggagaggtt gaggttcgct cccgaaaggt 240
aagacgagtc tgggggggaa atgatggggg tgtccggccc atagaggaca tccaggggtga 300
ctgggtcact gcggtttgca ctactgagt tctggattcc acatacatag gctcttgctg 360
catttcttgt gacattgaat agagttaggg tcctgttgcc attggacag 409

<210> 220
<211> 635
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (1)...(635)
<223> n=A,T,C or G

<400> 220
acagtgatag ctccccctgg gcaatacaat acaagaacag tgggttttgt caaattggaa 60
caaggaaaca gaaccacaga aataaatata ttgggttaaca tcagattagt tcaggttact 120
tttttgtaaa agttaagta gaggggactt ctgtattatg ctaactcaag tagactggaa 180
tctcctgtgt tctttttttt ttaaatttgt ttttaatttt ttttaattgga tctatcttct 240
tccttaacat ttcagttgga gtatgtagca ttttagcacca ctggetcaat gcgctcacct 300
aggtgagagn gngaccaaat cttaaagcat tagngctatt atcagttacc accatttggg 360
gcttttatcc ttcattgggt atgatgttct cctgatgaca catttctntg agttttgtaa 420
ttccagccaa agagagacca ttcactatth gatggctggc tgcattgana catttaaagc 480
ttttanagaa tacactacac cagggagtat gactactagt atgactatta ggagggtaat 540
accaagagtt ggactacgca ccttaggcaa gatncaaacc anctaaaata gaataaagaa 600
tgagtcagat gagtgtagcc attttaacca agcag 635

<210> 221
<211> 484
<212> DNA

<213> Homo sapiens

<400> 221

```
actccctggt ttgagaaact ttcttgaaga acaccatagc atgctgggtg tagttgggtgc 60
tcaccactcg gacgaggtaa ctctttaatc cagggttaact cttaatgttg cccagcgtga 120
actcgccggg ctggcaacct ggaacaaaag tcctgatcca gtagtcacac ttctttttcc 180
taaacaggac ggaggtgaca ttgtagctct tgtcttcttt cagctcatag atggtggcat 240
acatcttttg cgggtctttg tcttctctga gaattgcatt ccctgccagg cctaccacat 300
accacttccc ctggaattgg ttgtcctgga agttctgctg cagagggacc ttgctcagag 360
gtggggctgg gatcaggtct gaggtggagt cctgggctg ggcatgcaga gcccccaaca 420
gggctaggcc cagccacagg agacctaggg gcatgatttc agggccgagg aagcaggcgc 480
tgtg                                         484
```

<210> 222

<211> 566

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<222> (1)...(566)

<223> n=A,T,C or G

<400> 222

```
acattaaagt gtgatacttg gttttgaaaa cattcnaaca gtctctgtgg aaatctgaga 60
gaaattggcg gagagctgcc gtgggtgcatt cctcctgtag tgcttcaagc taatgcttca 120
tcctctctaa taacttttga tagacagggg ctagtcgcac agacctctgg gaagccctgg 180
aaaacgctga tgcttggttg aagatctcaa gcgcagagtc tgcaagttca tcccctcttt 240
cctgaggtct gttggctgga ggctgcagaa cattggtgat gacatggacc acgccatttg 300
tggccatgat gtcaggctcg gcaacaggct ccttggtgac actcaccaca ttgtttttca 360
agctgacttc cagcttgta ccttgagag actttagccg caccagggcc ccgatgcctc 420
cgctaaccag gatttcatca ccaatgtggt atttcaggat gttggcaagt tccttggcat 480
ctccaagag tctgctccgt tctcttggtg gcagggctcg gaaggcttca tttgtgggag 540
caaagactgt gtagacttcc tttccc                                         566
```

<210> 223

<211> 478

<212> DNA

<213> Homo sapiens

<400> 223

```
caggtaactta tttcaacaat tcttagagat gctagctagt gttgaagcta aaaatagctt 60
tatttatgct gaattgtgat ttttttatgc caaatttttt ttagttctaa tcattgatga 120
tagcttgga ataaataatt atgccatggc atttgacagt tcattattcc tataagaatt 180
aaattgagtt tagagagaat ggtgggtgtg agctgattat taacagttac tgaaatcaaa 240
tatttatgtg ttacattatt ccatttgtat tttagggttt cttttacatt ctttttatat 300
gcattctgac attacatatt ttttaagact atggaaataa tttaaagatt taagctctgg 360
tggatgatta tctgctaagt aagtctgaaa atgtaatatt ttgataatac tgtaatatatac 420
ctgtcacaca aatgcttttc taatgtttta accttgagta ttgcagttgc tgctttgt 478
```

<210> 224

<211> 323

<212> DNA

<213> Homo sapiens

<400> 224

```
acgggcaccg gcttcccta cagatggta cccacctgca agtggatggg gatctgcaac 60
ttcaatcaat caacttcac ggaggccagc ccctccggcc ccagggaccc ccgatgatgc 120
```


caccttgccc taccatggaa ggacccccaa ccttcaaccc gcctgtgcca tatttcggga 180
ggctgcaagg agggctcaca gctcgaagaa ccatcatcat caagggtat gtgcctccca 240
caggcaagag ctttgctatc aacttcaagg tgggctcctc aggggacata gctctgcaca 300
ttaatccccg catgggcaac ggt 323

<210> 225
<211> 147
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (1)...(147)
<223> n=A,T,C or G

<400> 225
ttggacttct agactcacct gttctcactc cctgnttnaa ttnaaccag ncatgcaatg 60
ccaaataata naattgctcc ctaccagctg aacagggagg agtctgtgca gtttctgaca 120
cttgttggtg aacatgggta aatacaa 147

<210> 226
<211> 104
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (1)...(104)
<223> n=A,T,C or G

<400> 226
nncaggnaca tgtgtgaaaa caatatgtga tactaccata gtgagccatg antntntaaa 60
aaaaaataa atgttttggg ggngatntgt attctccaac ttgg 104

<210> 227
<211> 491
<212> DNA
<213> Homo sapiens

<400> 227
acactgttgg tggtatatgg ggatgggggt ctcggttaatt ttgtttatta tttatgttta 60
ttattatggt ttatcattaa ttattcaata aatttttatt taaaaagtcg cctacttag 120
aaatcttctg tgggggtggg agggacaaaa gattacaaac caaaactcag gagatggtaa 180
cactggaatt gataaaatca cctgggatta gtcgtataac tctgaaccac caaacctctg 240
ctatcaagcc ttgctacagt catggctgtc cagaaagatt tacagttatt tttctgagaa 300
aggatccatg ggctttaaga acttcagaac tttaagaact tcagaagttc ttaagttgct 360
gaagctcaag taacgaagtt gaatgcaatc aaaaaaagaa taccagggag tcaaggcttg 420
agaggcacat tcttatccta aagtgactgc tcaaacctga cgagaccaag taaattactg 480
aagatacaaa g 491

<210> 228
<211> 328
<212> DNA
<213> Homo sapiens

<400> 228
actcagcgcc agcatcgccc cacttgatgt tggagggatc tcgctcctgg aagatgggtga 60
tgggatttcc attgatgaca agcttcccgt tctcagcctt gacggtgcca tgggaatttgc 120

```

catgggtgga atcatattgg aacatgtaaa ccatgtagtt gaggtcaatg aaggggtcat 180
tgatggcaac aatatccact ttaccagagt taaaagcagc cctggtgacc aggcgcccaa 240
tacgaccaa tccgttgact ccgaccttca ccttcccat ggtgtctgag cgatgtggct 300
cggctggcga cgcaaaagaa gatgcggc 328

```

```

<210> 229
<211> 689
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<222> (1)...(689)
<223> n=A,T,C or G

```

```

<400> 229
accacagcat catcccttgg tccagaatct actaccttcc acagcggccc aggtccact 60
gaaacaacac tcctacctga caacaccaca gcctcaggcc tccttgaagc atctacgcc 120
gtccacagca gcactggatc gccacacaca aactgtccc ctgccgntc tacaaccct 180
cagggagaat ctaccacctt ccagagctgg ccaaactcga aggacactac ccctgcacct 240
cctactacca catcagcctt tgttgagcta tctacaacct cccacggcag cccgagctca 300
actccaacaa cccacttttc tgccagctcc acaaccttgg gccgtagtga ggaatcgaca 360
acagtcacac gcagcccagt tgcaactgca acaacacct cgccgtccca ctccacaacc 420
tcaggcctcg ttgaagaatc tacgacctac cacagcagcc cgggctcaac tcaaacaatg 480
cacttccttg aaagcgacac aacttcaggc cgtggtgaag aatcaacaac ttcccacagc 540
agcacaacac acaaatatc ttcagctcct agcaccacat ctgcccttgt tgaagaacct 600
accagctacc acagcagccc gggctcaact gcaacaacac acttccttgc acaggttcca 660
caacctcaag gccgtagtgg agggaaatc 689

```

```

<210> 230
<211> 483
<212> DNA
<213> Homo sapiens

```

```

<400> 230
gggttctagc tcctccaatc ccattttatc ccatggaacc actaaaaaca aggtctgctc 60
tgctcctgaa gccctatatg ctggagatgg acaactcaat gaaaatttaa agggaaaacc 120
ctcaggcctg aggtgtgtgc cactcagaga cttcacctaa ctagagacag gcaaactgca 180
aaccatggtg agaaattgac gacttcacac tatggacagc ttttcccaag atgtcaaaac 240
aagactcctc atcatgataa ggctcttacc cccttttaat ttgtccttgc ttatgcctgc 300
ctctttcgct tggcaggatg atgctgtcat tagtatttca caagaagtag cttcagaggg 360
taacttaaca gagtgtcaga tctatcttgt caatcccaac gttttacata aaataagaga 420
tccttttagt caccagtgat ctgacattag cagcatcttt aacacagccg tgtgttcaaa 480
tgt 483

```

```

<210> 231
<211> 447
<212> DNA
<213> Homo sapiens

```

```

<400> 231
accctctcta ttcactagct tctgaaaagg gaggagtatt tttagtttga caatttaata 60
atttaaaaac aagacatctc caggtaggaa aaaatgaaag ctatttcatg caaacattat 120
ctaatttagc ttaaaagtga aagtggtaat actgttggtt tctgtaaatg ttgcagggtt 180
ttaaacttta taattacttt aatatttttg ataactagaa atctagtatt gccataaagg 240
aaactaagtg cccatcaaag atttgtttgg tataaataaa gaattatttg ttttgttttc 300
aatgacagta agctacaaat catgatgctt aaaaactttc taaagatgaa ttgtgtggca 360
gtgattggtc tgtttgtgga gaatgtatga aagctattaa tattctagaa tagattaata 420

```

aattggctat gttgttccaa tgaatgt

447

<210> 232

<211> 649

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<222> (1)...(649)

<223> n=A,T,C or G

<400> 232

```
gtgggcagaa gaaaaagcta gtgatcaaca gtggcaatgg agctgtggag gacagaaagc 60
caagtggact caacggagag gccagcaagt ctcaggaaat ggtgcatttg gtgaacaagg 120
agtcgtcaga aactccagac cagtttatga cagctgatga gacaagggaac ctgcagaatg 180
tggacatgaa gattgggggtg taacacctac accattatct tggaaagaaa caaccgttgg 240
aaacataacc attacaggga gctgggacac ttaacagatg caatgtgcta ctgattgttt 300
cattgcgaat ctttttttagc ataaaatttt ctattctttt tgttttttgt gttttgttct 360
ttaaagtcag gtccaatttg taaaaacagc attgctttct gaaattaggg cccaattaat 420
aatcagcaag aatttgatcg ttccagttcc cacttggagg cctttcatcc ctcggtgtg 480
ctatggatgg cttctaacaa aaactacaca tatgtattcc tgatcgccaa cctttcccc 540
accagctaag gacatttccc agggttaata gggcctggtc cctgggagga aatttgaatg 600
ggtccatttt gcccttncat agcctaattc ctgggcattg ctttncact 649
```

<210> 233

<211> 396

<212> DNA

<213> Homo sapiens

<400> 233

```
acaatgcaaa acataagtaa tcttttccact attataaacac ttgtatgatt ttaagacaaa 60
cttggcttaa attaagtttt ggggtcagcc ccaaattcct gcccttcac tgtattttga 120
attattttta aactctcaga tacagcttta tagttaaac attattagac tatatattct 180
aaattctaaa gtgaccaaag gggacagttt atgtaaagat aacacttttt cttaattttt 240
agaaaaccat tctttcatct cctgggtggc ttctttttcc gtctctatct cttttgttag 300
catcctatct ggtagtttgt taatatacat cttccctgag tgtttttaca acacaaagcc 360
atttagtgat tctgaatggc tactctgcct gccagt 396
```

<210> 234

<211> 4627

<212> DNA

<213> Homo sapiens

<400> 234

```
tcacttgcct gatattttcca gtgtcagagg gacacagcca acgtgggggc ccttctaggc 60
tgacagccgc tctccagcca ctgccgcgag cccgtctgct cccgccctgc ccgtgcactc 120
tccgcagccg ccctccgcca agccccagcg cccgtcccca tcgccgatga ccgcggggag 180
gaggatggag atgctctgtg ccggcagggt ccctgcgctg ctgctctgcc tgggtttcca 240
tcttctacag gcagtcctca gtacaactgt gattccatca tgtatcccag gagagtccag 300
tgataactgc acagcttttag ttcagacaga agacaatcca cgtgtggctc aagtgtcaat 360
aacaagtggt agctctgaca tgaatggcta ttgtttgcat ggacagtga tctatctggt 420
ggacatgagt caaaactact gcagggtgtg agtgggttat actggtgtcc gatgtgaaca 480
cttcttttta accgtccacc aacctttaag caaagagtat gtggctttga ccgtgattct 540
tattattttg tttcttatca cagtcgtcgg ttccacatat tatttctgca gatggtacag 600
aatcgaaaa agtaaagaac caaagaagga atatgagaga gttacctcag gggatccaga 660
```

| | | | | | | |
|-------------|-------------|-------------|-------------|-------------|-------------|------|
| gttgccgcaa | gtctgaatgg | cgccatcaaa | cttatgggca | gggataacag | tgtgcctggg | 720 |
| taataattaat | attccatttt | attaataata | tttatgttgg | gtcaagtgtt | aggtcaataa | 780 |
| cactgtattt | taatgtactt | gaaaaatgtt | tttatttttg | ttttattttt | gacagactat | 840 |
| ttgctaattgt | ataatgtgca | gaaaatattt | aatatcaaaa | gaaaattgat | atttttatac | 900 |
| aagtaatttc | ctgagctaaa | tgcttcattg | aaagcttcaa | agtttatatg | cctgggtgcac | 960 |
| agtgcctaga | agtaagcaat | tcccagggtca | tagctcaaga | attgttagca | aatgacagat | 1020 |
| ttctgtaagc | ctatatatat | agtcaaatcg | atttagtaag | tatgtttttt | atgttcctca | 1080 |
| aatcagtgat | aattggtttg | actgtaccat | ggtttgatat | gtagttggca | ccatgggtatc | 1140 |
| atataattaa | acaataatgc | aattagaatt | tgggagaagc | aaatataggt | cctgtgttaa | 1200 |
| acactacaca | tttgaaacaa | gctaaccctg | gggagtctat | ggctctctca | ctcagggtctc | 1260 |
| agctataatt | ctgttatatg | aggggcagtg | gacagttccc | tatgccaaact | cacgactcct | 1320 |
| acagggtacta | gtcactcatc | taccagattc | tgccatagta | aaatgaattg | aaaaacaatt | 1380 |
| ttctgtaatc | ttttatttta | gtagtgggca | tttcatagct | tcacaatgtt | ccttttttgt | 1440 |
| atattacaac | atttatgtga | ggtaattatt | gctcaacaga | caattagaaa | aaagtccaca | 1500 |
| cttgaagcct | aaatttgtgc | tttttaagaa | tatttttaga | ctatttcttt | ttataggggc | 1560 |
| tttgctgaat | tctaacatta | aatcacagcc | caaaatttga | tggactaatt | attattttta | 1620 |
| aatatatgaa | gacaataatt | ctacatgttg | tcttaagatg | gaaatacagt | tatttcatct | 1680 |
| tttattcaag | gaagttttaa | ctttaataca | gctcagtaaa | tggcttcttc | tagaatgtaa | 1740 |
| agttatgtat | ttaaagttgt | atcttgacac | aggaaatggg | aaaaaactta | aaaattaata | 1800 |
| tggtgtattt | ttccaaatga | aaaatctcaa | ttgaaagctt | ttaaaatgta | gaaacttaaa | 1860 |
| cacaccttcc | tgtggaggct | gagatgaaaa | ctagggctca | ttttcctgac | atttgtttat | 1920 |
| tttttggaag | agacaaagat | ttcttctgca | ctctgagccc | ataggtctca | gagagttaat | 1980 |
| aggagtattt | ttgggctatt | gcataaggag | ccactgctgc | caccactttt | ggattttatg | 2040 |
| ggaggctcct | tcatcgaatg | ctaaaccttt | gagtagagtc | tccctggatc | acataccagg | 2100 |
| tcaggaggga | tctgttcttc | ctctacgttt | atcctggcat | gtgctagggt | aaacgaaggc | 2160 |
| ataataagcc | atggctgacc | tctggagcac | cagggtgccag | gacttgtctc | catgtgtatc | 2220 |
| catgcattat | atacctgggt | gcaatcacac | gactgtcatc | taaagtcctg | gccctggccc | 2280 |
| ttactattag | gaaaataaac | agacaaaaac | aagtaaatat | atatggtcct | atacatattg | 2340 |
| tatatatatt | catatacaaa | catgtatgta | tacatgacct | taatggatca | tagaattgca | 2400 |
| gtcatttggg | gctctgctaa | ccatttatat | aaaacttaaa | aacaagagaa | aagaaaaatc | 2460 |
| aattagatct | aaacagttat | ttctgtttcc | tatttaatat | agctgaagtc | aaaatagta | 2520 |
| agaacacatt | ttaaatactc | tacttacagt | tggccctctg | tggttagttc | cacatctgtg | 2580 |
| gattcaacca | accaaggacg | gaaaatgctt | aaaaaataat | acaacaacaa | caaaaaatac | 2640 |
| attataacaa | ctatttactt | tttttttttt | cttttttgaga | tggagtctcg | ctctgttgcc | 2700 |
| caggttggag | tgcagtggca | cgatctcggc | tcactgcaac | ctcacctccc | gggttcaaga | 2760 |
| gatectcctg | cctcagcctc | ctgagcagct | gggactacag | gcgcatgcc | ccatgccag | 2820 |
| ctaatttttg | tatttttagt | agaggcgggg | tttcaccatg | ttggccagga | tggtctcaat | 2880 |
| ctcctaacct | tgagatccac | cctccacagc | ctcccaaact | gctgggatta | caggcgtgag | 2940 |
| ccaccgcacg | tagcatttac | attaggtatt | acaagtaatg | taaagatgat | ttaagtatac | 3000 |
| aggaggatgt | gaatagggtta | tatgcaagca | ctatgccctt | ttatataagt | gacttgaaca | 3060 |
| tctgtgcccc | atttttagtat | gtgcaggggg | gcgatctggg | aatcagtcct | ctgtggatac | 3120 |
| caagggtacaa | ctgtattttat | taacgccttac | tagatgtgag | gagagtctga | atattttcag | 3180 |
| tgatccttggc | tgtttcaaaa | aaatctattg | acttttcaat | aaatcagctg | caatccattt | 3240 |
| atttcattta | caaaagattt | attgtaagcc | tctcaatctt | ggtttttcag | ttgatcttaa | 3300 |
| gcatgtcaat | tcataaaaac | aagtcatttt | tgtatttttc | atctttaaga | atgcttaaaa | 3360 |
| aagctaattcc | ctaaaatagt | tagatctttg | taaatgcata | ttaaataata | aagtatgacc | 3420 |
| cacattactt | tttatgggtg | aaaataagac | aaaaataata | gttttagtga | ggatgggtgct | 3480 |
| gagtaaacad | aaaaactgat | ttgctctcag | ctgatgtgtc | ctgtacacag | tggaagatt | 3540 |
| ttagttcaca | cttagtctaa | ctccccatt | ttacagattt | ctcactatat | atattttctag | 3600 |
| aaggggctat | gcatattcaa | tgtattgaga | accaaagcaa | ccacaaatgc | ataaatgcat | 3660 |
| aatttatggg | cttcaaccaa | ggccacataa | taaccaggtt | aacttactct | ttaaccagga | 3720 |
| atattaaggt | ctataactag | tactcaaggt | ttaaccttaa | aattaagatt | tccttaacct | 3780 |
| taaccttaaa | attgatatta | tattaaacat | acataataca | atgtaactcc | actgttctcc | 3840 |
| tgaatatattt | ttgctctaata | ctctctgccg | aaagtcaaag | tgatgggaga | attgggtatac | 3900 |
| tggtatgact | acgtcttaag | tcagattttt | atttatgagt | ctttgagact | aaattcaatc | 3960 |
| accaccagggt | atcaaatcaa | cttttatgca | gcaaatatat | gattctagtg | tctgactttt | 4020 |
| gttaaatcca | gtaatgcagt | ttttaaaaac | ctgtatctga | cccactttgt | aatttttgct | 4080 |
| ccaatatcca | ttctgtagac | ttttgaaaaa | aaagttttta | atttgatgcc | caatatattc | 4140 |

tgaccgttaa aaaattcttg ttcatatggg agaaggggga gtaatgactt gtacaaacag 4200
 tattttctggt gtatatatta atgtttttaaa aaagagtaat ttcatttaaa tatctgttat 4260
 tcaaatttga tgatgttaaa tgtaataataa tgtattttct ttttattttg cactctgtaa 4320
 ttgcactttt taagtttgaa gagccatttt ggtaaacggt ttttattaaa gatgctatgg 4380
 aacataaagt tgtattgcat gcaatttaaa gtaacttatt tgactatgaa tattatcgga 4440
 ttactgaatt gtatcaattt gtttgtgttc aatatcagct ttgataattg tgtaccttaa 4500
 gatattgaag gagaaaatag ataatttaca agatattatt aatttttatt tatttttctt 4560
 gggaattgaa aaaaattgaa ataaataaaa atgcattgaa catcttgcac tcaaaatctt 4620
 cactgac 4627

<210> 235

<211> 169

<212> PRT

<213> Homo sapiens

<400> 235

Met Thr Ala Gly Arg Arg Met Glu Met Leu Cys Ala Gly Arg Val Pro
 5 10 15

Ala Leu Leu Leu Cys Leu Gly Phe His Leu Leu Gln Ala Val Leu Ser
 20 25 30

Thr Thr Val Ile Pro Ser Cys Ile Pro Gly Glu Ser Ser Asp Asn Cys
 35 40 45

Thr Ala Leu Val Gln Thr Glu Asp Asn Pro Arg Val Ala Gln Val Ser
 50 55 60

Ile Thr Lys Cys Ser Ser Asp Met Asn Gly Tyr Cys Leu His Gly Gln
 65 70 75 80

Cys Ile Tyr Leu Val Asp Met Ser Gln Asn Tyr Cys Arg Cys Glu Val
 85 90 95

Gly Tyr Thr Gly Val Arg Cys Glu His Phe Phe Leu Thr Val His Gln
 100 105 110

Pro Leu Ser Lys Glu Tyr Val Ala Leu Thr Val Ile Leu Ile Ile Leu
 115 120 125

Phe Leu Ile Thr Val Val Gly Ser Thr Tyr Tyr Phe Cys Arg Trp Tyr
 130 135 140

Arg Asn Arg Lys Ser Lys Glu Pro Lys Lys Glu Tyr Glu Arg Val Thr
 145 150 155 160

Ser Gly Asp Pro Glu Leu Pro Gln Val
 165

<210> 236

<211> 894

<212> DNA

<213> Homo sapiens

<400> 236

atgcacacc atcaccatca cacggccgcg tccgataact tccagctgtc ccagggtggg 60
 cagggatcgc ccattccgat cgggcaggcg atggcgatcg cgggccagat caagcttccc 120


```

accgttcata tcgggcctac cgccttcctc ggcttgggtg ttgtcgacaa caacggcaac 180
ggcgcacgag tccaacgcgt ggtcgggagc gctccggcgg caagtctcgg catctccacc 240
ggcgacgtga tcaccgcggt cgacggcgct ccgatcaact cggccaccgc gatggcggac 300
gcgcttaacg ggcatcatcc cggtgacgtc atctcgggtga cctggcaaac caagtcgggc 360
ggcacgcgta cagggaacgt gacattggcc gagggacccc cggccgaatt cgatgccttc 420
ctgaaatatg agaaggccga caaatactac tacacaagaa aatgtcgcaa tctgctgtcc 480
ttcctgaggg gcacctgctc attttgcagc cgcacactga gaaagcaatt ggatcacaac 540
ctcaccttcc acaagctggt ggcctatatg atctgcctac atacagctat tcacatcatt 600
gcaacacctgt ttaactttga ctgctatagc agaagccgac aggccacaga tggctccctt 660
gcctccattc tctccagcct atctcatgat gagaaaaagg ggggttcttg gctaaatccc 720
atccagtcctc gaaacacgac agtggagtat gtgacattca ccagccgggg tcaaacagag 780
gagagcatga atgagagtca tcctcgcaag tgtgcagagt cttttgagat gtgggatgat 840
cgtgactccc actgtaggcg ccctaagttt gaagggcatc cccctgagtc ttaa 894

```

<210> 237

<211> 297

<212> PRT

<213> Homo sapiens

<400> 237

```

Met His His His His His Thr Ala Ala Ser Asp Asn Phe Gln Leu
1      5      10     15
Ser Gln Gly Gly Gln Gly Phe Ala Ile Pro Ile Gly Gln Ala Met Ala
20     25     30
Ile Ala Gly Gln Ile Lys Leu Pro Thr Val His Ile Gly Pro Thr Ala
35     40     45
Phe Leu Gly Leu Gly Val Val Asp Asn Asn Gly Asn Gly Ala Arg Val
50     55     60
Gln Arg Val Val Gly Ser Ala Pro Ala Ala Ser Leu Gly Ile Ser Thr
65     70     75     80
Gly Asp Val Ile Thr Ala Val Asp Gly Ala Pro Ile Asn Ser Ala Thr
85     90     95
Ala Met Ala Asp Ala Leu Asn Gly His His Pro Gly Asp Val Ile Ser
100    105    110
Val Thr Trp Gln Thr Lys Ser Gly Gly Thr Arg Thr Gly Asn Val Thr
115    120    125
Leu Ala Glu Gly Pro Pro Ala Glu Phe Asp Ala Phe Leu Lys Tyr Glu
130    135    140
Lys Ala Asp Lys Tyr Tyr Tyr Thr Arg Lys Cys Arg Asn Leu Leu Ser
145    150    155    160
Phe Leu Arg Gly Thr Cys Ser Phe Cys Ser Arg Thr Leu Arg Lys Gln
165    170    175
Leu Asp His Asn Leu Thr Phe His Lys Leu Val Ala Tyr Met Ile Cys
180    185    190
Leu His Thr Ala Ile His Ile Ile Ala His Leu Phe Asn Phe Asp Cys
195    200    205
Tyr Ser Arg Ser Arg Gln Ala Thr Asp Gly Ser Leu Ala Ser Ile Leu
210    215    220
Ser Ser Leu Ser His Asp Glu Lys Lys Gly Gly Ser Trp Leu Asn Pro
225    230    235    240
Ile Gln Ser Arg Asn Thr Thr Val Glu Tyr Val Thr Phe Thr Ser Arg
245    250    255
Gly Gln Thr Glu Glu Ser Met Asn Glu Ser His Pro Arg Lys Cys Ala
260    265    270
Glu Ser Phe Glu Met Trp Asp Asp Arg Asp Ser His Cys Arg Arg Pro
275    280    285
Lys Phe Glu Gly His Pro Pro Glu Ser
290    295

```


<210> 238
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> PCR primer

<400> 238
ttttcttgtag tagtagtatt tgtcg 25

<210> 239
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> PCR primer

<400> 239
tgtcgcaatc tgctgtcctt cc 22

<210> 240
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> PCR primer

<400> 240
gctggtgaat gtcacatact cc 22

<210> 241
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> PCR primer

<400> 241
cggggtcaaa cagaggagag 20

<210> 242
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> PCR primer

<400> 242
gtcgaattcg atgccttcct gaaatatgag aag 33

<210> 243
<211> 33

<212> DNA

<213> Artificial Sequence

<220>

<223> PCR primer

<400> 243

cacctcgagt taagactcag ggggatgccc ttc

33

<210> 244

<211> 2609

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<222> (1)...(2609)

<223> n = A, T, C or G

<400> 244

| | | | | | | |
|-------------|------------|------------|------------|------------|-------------|------|
| gctgatagca | cagttctgtc | cagagaagga | aggcggaata | aacttattca | ttcccaggaa | 60 |
| ctcttggggt | aggtgtgtgt | ttttcacatc | ttaaaggctc | acagaccctg | cgctggacaa | 120 |
| atgttccatt | cctgaaggac | ctctccagaa | tccggattgc | tgaatcttcc | ctgttgccca | 180 |
| gaagggctcc | aaaccacctc | ttgacaatgg | gaaactgggt | ggttaaccac | tggttttcag | 240 |
| ttttgtttct | ggttggttgg | ttagggctga | atgttttctc | gtttgtggat | gccttcctga | 300 |
| aatatgagaa | ggccgacaaa | tactactaca | caagaaaaat | ccttgggtca | acattggcct | 360 |
| gtgcccagac | gtctgtcttc | tgcttgaatt | ttaacagcac | gctgactctg | cttcctgtgt | 420 |
| gtcgcaatct | gctgtccttc | ctgaggggca | cctgctcatt | ttgcagccgc | acactgagaa | 480 |
| agcaattgga | tcacaacctc | accttccaca | agctggtggc | ctatatgata | tgccctacata | 540 |
| cagctattca | catcattgca | cacctgttta | actttgactg | ctatagcaga | agccgacagg | 600 |
| ccacagatgg | ctcccttgcc | tccattctct | ccagcctatc | tcatgatgag | aaaaaggggg | 660 |
| gttcttggct | aaatcccctc | cagtcccga | acacgacagt | ggagtatgtg | acattcacca | 720 |
| gcgttgctgg | tctcactgga | gtgatcatga | caatagcctt | gattctcatg | gttacttcag | 780 |
| ctactgagtt | catccggagg | agttattttg | aagtcttctg | gtatactcac | caccttttta | 840 |
| tcttctatat | ccttggctta | gggattcacg | gcattggtgg | aattgtccgg | ggtcaaacag | 900 |
| aggagagcat | gaatgagagt | catcctcgca | agtgtgcaga | gtcttttgag | atgtgggatg | 960 |
| atcgtgactc | ccactgtagg | cgccctaagt | ttgaagggca | tccccctgag | tcttgggaagt | 1020 |
| ggatccttgc | accggtcatt | ctttatatct | gtgaaaggat | cctccgggtt | taccgctccc | 1080 |
| agcagaaggt | tgtgattacc | aagggtgtta | tgcacccatc | caaagttttg | gaattgcaga | 1140 |
| tgaacaagcg | tggcttcagc | atggaagtgg | ggcagtatat | ctttgttaat | tgcccctcaa | 1200 |
| tctctctcct | ggaatggcat | ccttttactt | tgacctctgc | tccagaggaa | gatttcttct | 1260 |
| ccattcataat | ccgagcagca | ggggactgga | cagaaaatct | cataagggct | ttcgaacaac | 1320 |
| aatattcacc | aattcccagg | attgaagtgg | atggtccctt | tggcacagcc | agtgaggatg | 1380 |
| ttttccagta | tgaagtggct | gtgctgggtg | gagcaggaat | tggggtcacc | ccctttgctt | 1440 |
| ctatcttgaa | atccatctgg | tacaaattcc | agtgtgcaga | ccacaacctc | aaaacaaaaa | 1500 |
| agatctatct | ctactggatc | tgcagggaga | caggtgcctt | ttcctgggtc | aacaacctgt | 1560 |
| tgacttccct | ggaacaggag | atggaggaat | taggcaaagt | gggttttcta | aactaccgtc | 1620 |
| tcttctctac | cggatgggac | agcaatattg | ttggtcatgc | agcattaaac | tttgacaagg | 1680 |
| ccactgacat | cgtgacaggt | ctgaaacaga | aaacctcctt | tgggagacca | atgtgggaca | 1740 |
| atgagttttc | tacaatagct | acctcccacc | ccaagtctgt | agtgggagtt | ttcttatgtg | 1800 |
| gccctcggac | tttggcaaag | agcctgcgca | aatgctgtca | ccgatattcc | agtctggatc | 1860 |
| ctagaaaggt | tcaattctac | ttcaacaaag | aaaatttttg | agttatagga | ataaggacgg | 1920 |
| taatctgcat | tttgtctctt | tgtatcttca | gtaattgagt | tataggaata | aggacggtaa | 1980 |
| tctgcatttt | gtctctttgt | atcttcagta | atttacttgg | tctcttcagg | tttgancagt | 2040 |
| cacttttagga | taagaatgtg | cctctcaagc | cttgactccc | tggatttctt | tttttgattg | 2100 |
| cattcaactt | cgttacttga | gcttcagcaa | cttaagaact | tctgaagttc | ttaaagttct | 2160 |
| gaanttctta | aagcccatgg | atcctttctc | agaaaaataa | ctgtaaatct | ttctggacag | 2220 |
| ccatgactgt | agcaaggctt | gatagcagaa | gtttggtggt | tcanaattat | acaactaatc | 2280 |

```

ccagggtgatt ttatcaattc cagtgttacc atctcctgag ttttggtttg taatcttttg 2340
tccctccac cccacagaa gattttaagt aggggtgactt tttaaataaa aatttattga 2400
ataattaatg ataaaacata ataataaaca taaataataa acaaaattac cgagaacccc 2460
atccccatat aacaccaaca gtgtacatgt ttactgtcac ttttgatag gtttatccag 2520
tgtgaacagc aatttattat ttttgctcat caaaaaataa aggatttttt ttcacttgaa 2580
aaaaaaaaa aaaaaaaaaa aaaaaaaaaa 2609

```

<210> 245

<211> 564

<212> PRT

<213> Homo sapiens

<400> 245

```

Met Gly Asn Trp Val Val Asn His Trp Phe Ser Val Leu Phe Leu Val
1      5      10      15
Val Trp Leu Gly Leu Asn Val Phe Leu Phe Val Asp Ala Phe Leu Lys
20     25     30
Tyr Glu Lys Ala Asp Lys Tyr Tyr Tyr Thr Arg Lys Ile Leu Gly Ser
35     40     45
Thr Leu Ala Cys Ala Arg Ala Ser Ala Leu Cys Leu Asn Phe Asn Ser
50     55     60
Thr Leu Ile Leu Leu Pro Val Cys Arg Asn Leu Leu Ser Phe Leu Arg
65     70     75     80
Gly Thr Cys Ser Phe Cys Ser Arg Thr Leu Arg Lys Gln Leu Asp His
85     90     95
Asn Leu Thr Phe His Lys Leu Val Ala Tyr Met Ile Cys Leu His Thr
100    105    110
Ala Ile His Ile Ile Ala His Leu Phe Asn Phe Asp Cys Tyr Ser Arg
115    120    125
Ser Arg Gln Ala Thr Asp Gly Ser Leu Ala Ser Ile Leu Ser Ser Leu
130    135    140
Ser His Asp Glu Lys Lys Gly Gly Ser Trp Leu Asn Pro Ile Gln Ser
145    150    155    160
Arg Asn Thr Thr Val Glu Tyr Val Thr Phe Thr Ser Val Ala Gly Leu
165    170    175
Thr Gly Val Ile Met Thr Ile Ala Leu Ile Leu Met Val Thr Ser Ala
180    185    190
Thr Glu Phe Ile Arg Arg Ser Tyr Phe Glu Val Phe Trp Tyr Thr His
195    200    205
His Leu Phe Ile Phe Tyr Ile Leu Gly Leu Gly Ile His Gly Ile Gly
210    215    220
Gly Ile Val Arg Gly Gln Thr Glu Glu Ser Met Asn Glu Ser His Pro
225    230    235    240
Arg Lys Cys Ala Glu Ser Phe Glu Met Trp Asp Asp Arg Asp Ser His
245    250    255
Cys Arg Arg Pro Lys Phe Glu Gly His Pro Pro Glu Ser Trp Lys Trp
260    265    270
Ile Leu Ala Pro Val Ile Leu Tyr Ile Cys Glu Arg Ile Leu Arg Phe
275    280    285
Tyr Arg Ser Gln Gln Lys Val Val Ile Thr Lys Val Val Met His Pro
290    295    300
Ser Lys Val Leu Glu Leu Gln Met Asn Lys Arg Gly Phe Ser Met Glu
305    310    315    320
Val Gly Gln Tyr Ile Phe Val Asn Cys Pro Ser Ile Ser Leu Leu Glu
325    330    335
Trp His Pro Phe Thr Leu Thr Ser Ala Pro Glu Glu Asp Phe Phe Ser
340    345    350
Ile His Ile Arg Ala Ala Gly Asp Trp Thr Glu Asn Leu Ile Arg Ala

```

[illegible]